

## Catálogo de Dataset no Repositório de Dados da Embrapa: a Experiência do Projeto-piloto de Implantação de Gestão de Dados de Pesquisa no Laboratório Multiusuário de Bioinformática da Embrapa



*Empresa Brasileira de Pesquisa Agropecuária  
Embrapa Informática Agropecuária  
Ministério da Agricultura, Pecuária e Abastecimento*

## **DOCUMENTOS 172**

# Catálogo de *Dataset* no Repositório de Dados da Embrapa: a Experiência do Projeto-piloto de Implantação de Gestão de Dados de Pesquisa no Laboratório Multiusuário de Bioinformática da Embrapa

*Marcia Izabel Fugisawa Souza*

*Marcos Cezar Visoli*

*Tércia Zavaglia Torres*

*Autores*

**Embrapa Informática Agropecuária**  
Campinas, SP  
2020

Exemplares desta publicação podem ser adquiridos na:

**Embrapa Informática Agropecuária**

Av. Dr. André Tosello, 209 - Cidade Universitária  
Campinas, SP, Brasil  
CEP. 13083-886  
Fone: (19) 3211-5700  
www.embrapa.br  
www.embrapa.br/fale-conosco/sac

Comitê Local de Publicações  
da Unidade Responsável

Presidente

*Stanley Robson de Medeiros Oliveira*

Secretária-Executiva

*Maria Fernanda Moura*

Membros

*Adriana Farah Gonzalez, membro nato, Alexandre de Castro, membro indicado, Carla Cristiane Osawa, membro nato, Debora Pignatari Drucker, membro eleito, Ivan Mazoni, membro eleito, João Camargo Neto, membro indicado, João Francisco Gonçalves Antunes, membro eleito, Magda Cruciol, membro nato.*

Revisão de texto

*Adriana Farah Gonzalez*

Normalização bibliográfica

*Carla Cristiane Osawa*

Projeto gráfico da coleção

*Carlos Eduardo Felice Barbeiro*

Editoração eletrônica

*Marília Bastos sob supervisão de Magda Cruciol*

Arte da capa

*Marília Bastos*

**1ª edição**

Versão digital (2020)

**Todos os direitos reservados.**

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei nº 9.610).

**Dados Internacionais de Catalogação na Publicação (CIP)**

Embrapa Informática Agropecuária

---

Souza, Márcia Izabel Fugisawa.

Catalogação de dataset no Repositório de Dados da Embrapa : a experiência do Projeto-piloto de implantação de Gestão de Dados de Pesquisa no Laboratório Multiusuário de Bioinformática / Márcia Izabel Fugisawa Souza, Marcos Cezar Visoli, Tércia Zavaglia Torres. – Campinas : Embrapa Informática Agropecuária, 2020.

PDF (117 p.): il. color. - (Documentos / Embrapa Informática Agropecuária, ISSN 1677-9274 ; 172).

1. Representação descritiva. 2. Representação temática. 3. Catalogação. 4. Dataset. 5. Repositório de Dados da Embrapa. 6. AgDados. I. Visoli, Marcos Cezar. II. Torres, Tércia Zavaglia. III. Título. IV. Embrapa Informática Agropecuária. V. Série.

CDD (21. ed.) 025.32

## Autores

### **Marcia Izabel Fugisawa Souza**

Bacharel em Biblioteconomia, Doutora em Educação, Analista da Embrapa Informática Agropecuária, Campinas, SP.

### **Marcos Cezar Visoli**

Cientista da Computação, Mestre em Informatique e Systèmes Spécialité Recherche Modèles, Systèmes, Imagerie, Robotique, Pesquisador da Embrapa Informática Agropecuária, Campinas, SP.

### **Tércia Zavaglia Torres**

Administradora, Doutora em Educação, Analista da Embrapa Informática Agropecuária, Campinas, SP.

## Agradecimentos

Os autores manifestam reconhecimento e gratidão aos pesquisadores Antonio Nhani Júnior, Paula Regina Kuser Falcão e Poliana Fernanda Giachetto, e aos analistas Leandro Carrijo Cintra e Luiz Falaguasta Barbosa, pelas interações e contribuições em todas as etapas de estabelecimento do processo catalogação no Projeto-piloto de Implantação de Gestão de Dados de Pesquisa no Laboratório Multiusuário de Bioinformática da Embrapa (LMB). Essa colaboração e acompanhamento permanentes foram imprescindíveis para que o trabalho desenvolvido no processo de catalogação pudesse refletir as especificidades, necessidades e interesses do LMB.

Os autores agradecem as colegas Adriana Delfino dos Santos, Daniela Maciel Pinto, Vera Viana dos Santos Brandão e Carla Cristiane Osawa pelo acurado trabalho de revisão e pelas contribuições para a melhoria deste documento.

## Apresentação

O modus operandi de fazer ciência vem, ao longo dos tempos, se modificando. Esta transformação perceptível se traduz em avanços na produção de novas informações e conhecimento que, por conseguinte, se materializam na geração de tecnologias, produtos e serviços que atendem em amplo espectro a diversa gama de interesses da sociedade. No segmento da agropecuária brasileira esta é uma realidade protagonizada pela Embrapa há 47 anos.

É notório que vimos passando de um paradigma de ciência baseada em hipótese para um paradigma de ciência orientada a dados. Por detrás deste último paradigma encontram-se concepções inter e multidisciplinares advindas de vários domínios de conhecimento, fortemente beneficiadas pelos impactos promovidos pelas tecnologias de informação e comunicação (TIC). O paradigma de ciência orientada a dados traz implicações substanciais nas práticas de pesquisas em instituições de pesquisa, desenvolvimento e inovação (PD&I), como a Embrapa.

No cerne deste paradigma observa-se a potência das TIC, ao servirem como ferramentas para favorecer, facilitar e, principalmente, ampliar a capacidade de os pesquisadores desenvolverem um fazer científico baseado no compartilhamento e no reuso de dados, uso de diferentes tipos dados de pesquisa científica e de infraestrutura tecnológica que possibilita práticas colaborativas de pesquisa entre os cientistas.

Atenta à essas transformações, a Embrapa instituiu em 2019 uma Política de Governança de Dados, Informação e Conhecimento (PGDIC), cujo objetivo é estabelecer princípios, diretrizes, atribuições e responsabilidades para a gestão de dados, informação e conhecimento, bem como promover a divulgação de informações relevantes na Empresa para fornecer mecanismos de geração, organização, tratamento, acesso, preservação, recuperação, divulgação, compartilhamento e reuso dos seus ativos de informação.

Esse é o contexto em que a presente publicação está inserida. Ademais, cabe destacar que se trata de um dos vários resultados obtidos com a realização de uma Ação Gerencial Local (AGL) intitulada “Implantar a gestão de dados de pesquisa na Embrapa Informática Agropecuária, em alinhamento ao Programa de Governança de Dados de Pesquisa”, desenvolvida na Embrapa Informática Agropecuária. Esta AGL integra uma Ação Gerencial Corporativa liderada pela Secretaria de Desenvolvimento Institucional da Embrapa. Esta AGL teve o objetivo de desenhar, sob a perspectiva da metodologia de governança de processos da Embrapa, o processo de gestão de dados de pesquisa que ocorre no Laboratório Multiusuário de Bioinformática da Embrapa (LMB), e ao final propor melhorias.

Esta publicação é o registro de parte das melhorias propostas ao processo de gestão de dados do LMB. Contém as regras básicas para a catalogação de datasets

gerados pelos projetos de pesquisa liderados pelo LMB. Tem o objetivo de orientar o trabalho do catalogador no uso do Repositório de Dados da Embrapa (AgDados). Trata-se, portanto, de uma importante contribuição para a implementação da PGDIC, visto que a catalogação é uma atividade essencial para a interoperabilidade, o compartilhamento, o uso e o reúso dos dados de pesquisa.

*Silvia Maria Fonseca Silveira Massruhá*

Chefe-geral





## Sumário

<b>Introdução</b>	<b>13</b>
<b>1. Descrição de Metadados – introdução</b>	<b>15</b>
Metadados e o Software Dataverse: a Experiência da Embrapa Informática Agropecuária	18
Catalogação de dados	18
Catalogação de Dados no Software Dataverse: a Experiência do Projeto-piloto Implantação de GDP Laboratório Multiusuário de Bioinformática da Embrapa (LMB)	19
<b>2. Regras Gerais para Descrição de <i>Dataset</i> no Repositório de Dados da Embrapa (AgDados)</b>	<b>20</b>
Elementos metadados requeridos, fortemente recomendados, recomendados e opcionais	21
<b>3. Descrição dos Metadados de Citação</b>	<b>24</b>
Título*	25
Subtítulo	26
Título alternativo	26
URL alternativa	26
Outro identificador	27
Agência	27
Identificador	27
Autor	28
Nome*	28
Afiliação	28
Sistema do identificador	29
Identificador	29
Contato	29
Nome	30
Afiliação	30
E-mail do contato*	30
Descrição*	31
Texto*	31
Data	31
Assunto*	32
Palavras-chave	32

Termo-chave	32
Vocabulário controlado do termo-chave	33
URL do termo-chave	33
Classificação	33
Termo de classificação	34
Vocabulário controlado da classificação	34
URL do termo de classificação	34
Publicação relacionada	35
Citação	35
Tipo de identificador	35
Número do identificador	36
URL da publicação	36
Notas	36
Idioma	37
Produtor	37
Nome do produtor	37
Afiliação do produtor	38
Abreviatura do nome do produtor	38
URL do produtor	38
URL do logotipo do produtor	39
Data de produção	39
Local de produção	39
Colaborador	39
Tipo de colaborador	40
Nome do colaborador	40
Financiamento	40
Nome do órgão financiador	41
Número do contrato	41
Distribuidor	41
Nome do distribuidor	42
Afiliação do distribuidor	42
Abreviatura do nome do distribuidor	42
URL do distribuidor	43
URL do logotipo do distribuidor	43

Data de distribuição	43
Depositante*	44
Data do depósito	44
Período da cobertura de dados	44
Data de início	44
Data de término	45
Período da coleta de dados	45
Data de início	45
Data de término	46
Tipo de dados	46
Série	46
Nome da série	47
Informação sobre série	47
Software	47
Nome do software	47
Versão do software	48
Material relacionado	48
<i>Datasets</i> relacionados	48
Outras referências	49
Fontes de dados	49
Origem das fontes de dados	49
Características das fontes de dados observadas	50
Documentação e acesso a fontes de dados	50

<b>4. Descrição de Metadados <i>Life Science</i></b>	<b>50</b>
Tipo de design	51
Tipo de fator	52
Organismo	52
Outro organismo	52
Tipo de medida	52
Outro tipo de medida	53
Tipo de tecnologia	53
Plataforma tecnológica	53
Tipo de célula	53

<b>Referências</b>	<b>54</b>
<b>Apêndice A - Vocabulários Controlados</b>	<b>56</b>
Introdução	56
AGROVOC Thesaurus	58
O que é AGROVOC Thesaurus?	58
AGROVOC Thesaurus para Palavras-chave	58
AGROVOC Thesaurus para Classificação	63
NAL Thesaurus	67
O que é NAL Thesaurus?	67
NAL Thesaurus para Palavras-chave	67
NAL Thesaurus para Classificação	71
NCBI Taxonomy	76
O que é NCBI Taxonomy?	76
NCBI Taxonomy para Palavras-chave	76
NCBI Taxonomy para Classificação	79
<b>Referências</b>	<b>83</b>
<b>Apêndice B - Catalogação de <i>Dataset</i>: Guia Resumido</b>	<b>84</b>
Introdução	84
Descrição de Metadados de Citação	86
Descrição de Metadados <i>Life Science</i>	93
<b>Referências</b>	<b>93</b>
<b>Apêndice C – Revisão na Catalogação de <i>Dataset</i></b>	<b>95</b>
Introdução	95
Catalogação de <i>dataset</i> no AgDados	95
Por que revisar a catalogação de dataset no AgDados?	96
Controle de qualidade e garantia de qualidade	96
Revisão de catalogação de <i>dataset</i> no AgDados	97
Atividades de revisão na catalogação de <i>dataset</i>	98
Referências	105
<b>Anexo 1 – Modelo de Relatório da Revisão</b>	<b>107</b>

## Introdução

O presente documento reúne um conjunto de regras básicas para catalogação de *datasets*<sup>1</sup>, elaboradas com o objetivo de orientar o trabalho do catalogador no Repositório de Dados da Embrapa (AgDados). Essas regras de catalogação foram discutidas e definidas a partir de uma experiência conduzida no Projeto-piloto de Implantação de Gestão de Dados de Pesquisa no Laboratório Multiusuário de Bioinformática da Embrapa (LMB). Trata-se, portanto, de uma contribuição técnica orientada ao estabelecimento do processo de catalogação, compreendendo as atividades de representação descritiva e temática de *datasets*, como será visto adiante.

Por que catalogar *datasets*? Uma resposta possível para esta pergunta pode ser construída a partir da análise do paradigma da *e-Science* e da concepção de ciência aberta e colaborativa, esta entendida como um empreendimento multi e interdisciplinar da ciência contemporânea (Chan et al., 2015). Compreende-se, pois, essa ciência como aquela que tem “capacidade de gerar e armazenar dados em escala sem precedentes muito além da capacidade humana de análise” (Cesar Junior, 2011, p. 7-8).

Atenta a esse paradigma, a Empresa Brasileira de Pesquisa Agropecuária (Embrapa) tem se posicionado ao lado das instituições líderes mundiais, a respeito da adoção de políticas de governança de dados de pesquisa. No momento, tem como principal ação estratégica, a criação de mais um repositório de dados para abrigar o depósito e a preservação de dados de pesquisa, com vista a promover o reúso dos dados por ela produzidos. Portanto, trata-se de uma iniciativa que se soma às experiências já consolidadas de criação de repositórios na Empresa, como é o caso do Repositório Acesso Livre à Informação Científica da Embrapa (Alice)<sup>2</sup>, do Repositório de Informação Tecnológica da Embrapa (Infoteca-e)<sup>3</sup> e do Geoinfo<sup>4</sup> - Infraestrutura de Dados Espaciais da Embrapa.

Dados de pesquisa produzidos pela Embrapa são ativos digitais de valor intangível, com potencial para beneficiar diretamente inúmeros pesquisadores e instituições por meio do reúso, seja pelo reaproveitamento, agregação, integração, metanálise ou reanálise (Curty, 2019). No entanto, a implantação de repositórios de dados traz consigo a necessidade de que os dados sejam descritos e preservados adequadamente, de acordo com requisitos de

---

<sup>1</sup> Neste documento, optou-se pelo uso do termo *dataset* para denominar o recurso digital objeto da catalogação, ao invés de conjunto de dados.

<sup>2</sup> Disponível em: <<http://www.alice.cnptia.embrapa.br>>

<sup>3</sup> Disponível em: <<http://www.infoteca.cnptia.embrapa.br>>

<sup>4</sup> Disponível em: <<http://www.embrapa.br/geoinfo>>

qualidade para o reuso eficiente, e que atendam aos princípios FAIR<sup>5</sup>, e às orientações da biblioteconomia e da catalogação de dados (Semeler; Pinto, 2019).

Nesse contexto se insere a necessidade de implementação de melhorias de natureza tecnológica e processual, especialmente no que tange ao tratamento descritivo dos *datasets*, abarcado pela atividade de catalogação, que visa conferir qualidade tanto aos metadados quanto aos dados (*datasets*) em si.

Ademais, o processo de catalogação impacta diretamente outros processos dele decorrentes, como o da indexação, da busca, da descoberta e o da recuperação de dados e informações, para os quais os vocabulários controlados tornam-se imprescindíveis e de uso obrigatório. Diante da relevância de tais instrumentos para o tratamento de assuntos de *datasets*, foi preparado um documento de apoio ao uso de três vocabulários controlados, selecionados pelos pesquisadores do referido Projeto-piloto, para representação temática de dados biológicos, por meio da atribuição de palavras-chave (Apêndice A).

O presente documento está constituído em quatro capítulos e três apêndices. O primeiro capítulo salienta brevemente a importância da descrição de metadados e enfatiza a necessidade de atendimento aos princípios FAIR. A catalogação de *datasets* é apresentada a partir da experiência conduzida no Projeto-piloto Implantação de GDP Laboratório Multiusuário de Bioinformática da Embrapa (LMB), e com a utilização do software Dataverse.

O segundo capítulo traz as regras gerais que devem ser observadas durante a etapa de descrição de *dataset* no Repositório de Dados da Embrapa (AgDados). O terceiro capítulo é constituído das orientações para a descrição dos Metadados de Citação, considerados default do software Dataverse, os quais compõem um conjunto genérico de elementos, campos e subcampos. No quarto capítulo são apresentadas as regras para descrever Metadados Life Science, também oferecidos pelo Dataverse; neste caso, constituídos por elementos específicos voltados para a descrição de dados biológicos.

Ao final, encontram-se três Apêndices que foram preparados com o intuito facilitar e ilustrar o trabalho do catalogador. O Apêndice A orienta o uso de vocabulários controlados na atribuição de palavras-chave e de categorias de assunto para os *datasets*. Por sua vez, o Apêndice B é uma versão resumida do documento completo de catalogação, objeto principal desta publicação. O Apêndice C apresenta as orientações para a execução da atividade de revisão da catalogação de *dataset* no AgDados. Pretendeu-se com esses apêndices facilitar a tarefa

---

<sup>5</sup> FAIR é o acrônimo de *Findable, Accessible, Interoperable, Reusable*, princípios que preconizam que os dados necessitam ser encontráveis, acessíveis, interoperáveis e reutilizáveis (Go Fair, 2016).

de catalogação, bem como enriquecer a experiência de catalogador de *datasets* no Repositório de Dados da Embrapa (AgDados).

## Catálogo de *Dataset* no Repositório de Dados da Embrapa: a Experiência do Projeto-piloto de Implantação de Gestão de Dados de Pesquisa no Laboratório Multiusuário de Bioinformática da Embrapa

### 1. Descrição de Metadados – introdução

Este capítulo reúne as orientações para a catalogação de *datasets* no Repositório de Dados da Embrapa (AgDados). Essas orientações se resumem às regras gerais de descrição de *dataset* aplicáveis às especificidades do software Dataverse. Em especial, o documento traz orientações para os usuários (produtor de *dataset*, especialista do domínio, especialista no tratamento da informação, curador etc.) na realização da atividade de representação descritiva e de representação temática de *datasets*.

As orientações foram elaboradas com base em recomendações técnicas e nas regras básicas do Código de Catalogação Anglo-Americano - AACR2 (Ribeiro, 1995, 2003; Código..., 2004), bem como no uso de vocabulários controlados para tratamento de assunto (palavras-chave e classificação taxonômica), para padronização de nomes pessoais e corporativos, e-mails, afiliação etc.

O conteúdo deste documento orientador deverá ser submetido à constante avaliação e adequação, procurando atender às necessidades dos usuários do Projeto-piloto e ao aprimoramento das recomendações e orientações em relação a metadados e catalogação de *datasets* no AgDados.

Os metadados e a catalogação de dados no escopo deste documento devem ser considerados na perspectiva da gestão de dados de pesquisa e do ciclo de vida dos dados. Assim, metadados e catalogação de *datasets* correspondem a duas subetapas importantes da etapa Descrição, do ciclo de vida dos dados, e como tais devem estar alinhadas aos princípios FAIR (Go Fair, 2016), um acrônimo formado pelas palavras em inglês: *Findable*, *Accessible*, *Interoperable*, *Reusable*. Esses quatro princípios são compostos de 15 elementos a eles associados, e que contribuem para complementar e enriquecer as informações essenciais, tanto para metadados como para dados em si, objetos da catalogação. A seguir são apresentados os 15 elementos associados a cada princípio FAIR (Henning et al., 2019, p. 178-179).

<b>Princípio F - <i>Findable</i> (Encontrabilidade) e seus elementos</b>
F1. Os (meta)dados devem ter identificadores globais, persistentes e identificáveis.
F2. Os dados devem ser descritos com metadados enriquecidos (impacta diretamente R1).
F3. Os metadados devem incluir claramente e explicitamente os identificadores dos dados que descrevem.
F4. Os (meta)dados devem ser recuperáveis ou indexados em recursos que ofereçam capacidade de busca.

<b>Princípio A - <i>Accessible</i> (Acessibilidade) e seus elementos</b>
A1. (Meta)dados devem ser recuperáveis pelos seus identificadores usando protocolo de comunicação padronizado.
A1.1. O protocolo deve ser aberto, gratuito e universalmente implementável.
A1.2. O protocolo deve permitir procedimentos de autenticação e autorização, quando necessário.
A2. Metadados devem ser acessíveis, mesmo quando os dados não estão mais disponíveis.

<b>Princípio I - <i>Interoperable</i> (Interoperabilidade) e seus elementos</b>
I1. (Meta)dados devem ser representados por meio de uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento.
I2. (Meta)dados devem usar vocabulários que assegurem os princípios FAIR.
I3. (Meta)dados devem incluir referências qualificadas para outros (meta)dados.

<b>Princípio R - <i>Reusable</i> (Reusabilidade) e seus elementos</b>
R1. (Meta)dados são descritos com uma pluralidade de atributos precisos e relevantes.
R1.1. (Meta)dados devem ser disponibilizados com licenças de uso claras e acessíveis.
R1.1. (Meta)dados devem estar associados à sua proveniência.
R1.3. (Meta)dados devem estar alinhados com padrões relevantes ao seu domínio.

Cabe esclarecer que o detalhamento desses princípios não é objeto deste documento orientador de catalogação. Contudo, é oportuno enfatizar que os metadados e a catalogação são imprescindíveis para que dados de pesquisa possam ser “[...] identificáveis, citáveis, visíveis, recuperáveis, interpretáveis, contextualizáveis, interoperáveis e reutilizáveis onde quesitos de consistência e procedência são considerados” (Semeler; Pinto, 2019, p. 116).

Metadados possibilitam a exploração de outras dimensões e facetas do dado, que ao serem reveladas pela catalogação passam a contribuir para a melhoria da gestão e qualidade dos dados de pesquisa da Embrapa, favorecendo a descoberta das coleções de dados para



a comunidade científica. Metadados são indispensáveis para que, no futuro, os conteúdos digitais possam ser acessados e interpretados. Sem metadados, de acordo com Gray (2002), citado por Sayão e Sales (2016, slide 83), os usuários

[...] não saberão os detalhes de como os dados foram obtidos e preparados: 1) como os instrumentos foram projetados e construídos; 2) quando, onde e como os dados foram coletados; e, 3) não terão uma descrição dos processos que levaram aos dados derivados, que são tipicamente usados para análises científicas.

Metadados também são indispensáveis à interoperabilidade técnica e semântica, ou seja, sem eles os repositórios e plataformas de dados não poderão intercambiar dados e informações. Metadados são constituídos por elementos descritivos bem definidos, como por exemplo: autor, título, descrição, assunto, palavra-chave, identificador, produtor, tipo de dados, restrições de acesso, termo de uso das coleções etc., formando a partir da catalogação de dados um corpo de informações capaz de contextualizar os dados, sobretudo, quanto à proveniência, história, natureza e propósito.

A geração de metadados enriquecidos traz benefícios diretos para a gestão de dados, impactando positivamente no arquivamento e preservação, bem como na interoperabilidade e recuperação de conjuntos de dados de pesquisa. Dados somente serão úteis para análise se tiverem sido descritos com metadados de qualidade, e para que isso aconteça, a melhor recomendação é a adoção dos princípios FAIR para orientar a catalogação.

De acordo com Veiga (2019, p. 18-22), os elementos-chave que devem orientar a adoção dos princípios FAIR, no que diz respeito ao aspecto **descritivo** dos metadados, são:

- Elementos metadados para identificadores únicos e persistentes tanto para dados, como para o conjunto de dados.
- Conjunto de dados utilizando metadados enriquecidos com pluralidade de atributos precisos e relevantes.
- Elemento metadado que indique clara e explicitamente os identificadores persistentes, tanto do conjunto de dados como também do próprio metadado nos repositórios e plataformas de dados.
- Metadados registrados ou indexados em recursos de identificação que ofereçam capacidade de busca.
- Metadados utilizando protocolos padronizados de comunicação para facilitar a recuperação dos dados via metadados, inclusive.
- Disponibilidade de metadados para acesso, mesmo que os dados não estejam acessíveis e disponíveis.

- Elemento metadado para representação do conhecimento por meio de linguagem formal, e pelo uso de taxonomias e vocabulários controlados de acordo com os princípios FAIR, especializados e padronizados por área específica do domínio.
- Elemento metadado para referências qualificadas de conjunto de dados e outros objetos de pesquisa derivados, que se interconectam, assegurando interligações semânticas entre eles, e que sejam *linkáveis* para outros conjuntos de dados.
- Metadados com riqueza de atributos com alto nível de detalhes para que permitam ao pesquisador avaliar a possibilidade de reuso e adequação às suas necessidades.
- Elemento metadado com informações inequívocas, definindo claramente quem podem ter acesso aos dados, com que finalidade e sob quais condições.
- Elemento metadado que especifique a proveniência dos dados, o que subsidiará o pesquisador ao decidir sobre a utilidade dos dados ou metadados, e ao atribuir crédito ao produtor dos dados.
- Adoção de metadados deve estar alinhada com padrões relevantes e específicos da comunidade e da área de pesquisa.

## Metadados e o Software Dataverse: a Experiência da Embrapa Informática Agropecuária

O software Dataverse é uma aplicação web, de código-aberto, dedicada para o compartilhamento, preservação, citação, exploração e análise de dados de pesquisa. Os conjuntos de dados, também denominados *datasets*, armazenados no Dataverse, são compostos por metadados e os arquivos de dados. (Dataverse, 2020). O software Dataverse foi selecionado para apoiar na gestão de dados de pesquisa da Embrapa, considerando-se as experiências internacionais, as características tecnológicas, os recursos e as iniciativas existentes no desenvolvimento e implantação de tecnologias voltadas para este domínio. Dois projetos de pesquisa relacionados ao Laboratório Multiusuário de Bioinformática da Embrapa foram escolhidos como projetos-piloto com o objetivo de avaliar o uso do software Dataverse para a gestão de dados de pesquisa. Neste contexto, foram realizadas experiências pioneiras com metadados e catalogação de *datasets* na Embrapa, suportadas pelo software Dataverse.

## Catalogação de dados

A catalogação é uma disciplina da Biblioteconomia e da Ciência da Informação, e como tal requer: o emprego de formato e padrões internacionais (de metadados, inclusive) e de normas técnicas para a executar representação descritiva; e a adoção de taxonomias, tesouros e vocabulários controlados para formular a representação temática. Isso se aplica a

quaisquer artefatos, objetos e unidades de informação, sejam impressos ou digitais. Portanto, a catalogação vai além da simples inserção de objetos de dados e de informação em base de dados, repositórios ou plataforma de dados.

No contexto da Ciência Aberta e da Biblioteconomia de Dados, a catalogação de *datasets* deve ser implementada em conformidade com os princípios FAIR, para que metadados e dados sejam descritos com padrão de qualidade visando o reúso desses dados para a pesquisa. Ademais, a catalogação de dados beneficia a transparência da pesquisa, o que favorece a descoberta de dados. Também cria condições para a reprodutibilidade e verificação da pesquisa publicada. (Gordana; Dragan, 2017).

## Catálogo de Dados no Software Dataverse: a Experiência do Projeto-piloto Implantação de GDP Laboratório Multiusuário de Bioinformática da Embrapa (LMB)

O processo de catalogação de *datasets* se desenrola no ambiente do software *Dataverse*, utilizando-se um formulário para descrição dos metadados e upload de arquivos de dados. O software *Dataverse* oferece ao usuário a possibilidade de escolher qual conjunto de metadados deseja usar para melhor atender aos interesses e o contexto dos dados a serem descritos. O conjunto de metadados Citation Metadata, doravante denominado Metadados de Citação, é default do software *Dataverse*, sendo exportável para o padrão Dublin Core. Por conseguinte, este é o conjunto de metadados adotado no Projeto-piloto do LMB para a catalogação de *datasets*. Outro destacado conjunto de metadados é o Life Science Metadata, que após análise pela equipe do Projeto-piloto também foi escolhido para catalogar *datasets*, em razão da sua especificidade temática (ciências da vida), considerada adequada aos projetos de pesquisa do LMB. Desse modo, Metadados de Citação e Metadados Life Science são utilizados concomitantemente, visto que o software *Dataverse* possibilita o uso dos dois conjuntos de metadados no mesmo formulário de catalogação de *datasets*.

No projeto-piloto, o processo de catalogação dos *datasets* inicia-se com a atividade de **auto-depósito**, a qual deve ser executada pelo autor do dataset ou outra pessoa sob sua responsabilidade. O momento do auto-depósito do *dataverse* ou *dataset* no software *Dataverse* corresponde à fase de **pré-catalogação**, na qual são preenchidos os campos correspondentes aos elementos metadados básicos, sendo obrigatório informar: título, autor, contato, descrição, assunto e depositante.

Na fase seguinte ocorre a **catalogação** propriamente dita, começando pela revisão do preenchimento prévio dos conteúdos relativos aos elementos metadados básicos. A descrição catalográfica de *dataverses* e *datasets* se ocupa essencialmente do preenchimento dos

elementos metadados complementares, a ser executado pelo autor do *dataset*, na função de catalogador, exigindo deste um conhecimento mínimo de normas e padrões oriundos da Biblioteconomia e da Ciência da Informação.

O emprego de técnicas de catalogação para descrever e representar dados ou quaisquer objetos digitais de pesquisa, tomando por base um conjunto estruturado de elementos metadados FAIR, é essencial para garantir a interoperabilidade técnica e semântica, o compartilhamento, o uso e o reuso dos dados de pesquisa. Ao bibliotecário e/ou cientista da informação cabe a responsabilidade técnica por esta atividade.

## 2. Regras Gerais para Descrição de *Dataset* no Repositório de Dados da Embrapa (AgDados)

Nesta seção são apresentadas as regras que devem ser observadas na descrição de *dataset* no Repositório de Dados da Embrapa (AgDados). Essas regras orientam o preenchimento do formulário de catalogação, visando à clareza e padronização das informações. Todos os elementos metadados, campos e subcampos exigem atenção no seu preenchimento a fim de evitar erros de qualquer natureza. Ou seja, o preenchimento adequado dos campos e subcampos tem impacto direto sobre a qualidade dos *datasets* e de seus metadados, os quais são indexados e recuperados por mecanismos de busca tanto do software Dataverse como por quaisquer outros buscadores na Internet.

As regras básicas que orientam o preenchimento do formulário de catalogação são referentes aos itens:

- Elementos metadados obrigatório, recomendado e opcional.
- Pontuação.
- Uso de caracteres em maiúscula.
- Tecla Enter.
- Ajuda rápida.
- Nome de autor pessoal e nome autor corporativo.
- E-mail pessoal.
- E-mail corporativo.
- Afiliação.
- Palavras-chave e Termos livres.
- Data.

Em relação à catalogação de Metadados devem ser observadas as regras de exigência para preenchimento descritas a seguir.

## Elementos metadados requeridos, fortemente recomendados, recomendados e opcionais

No tocante ao preenchimento do formulário on-line de catalogação de *dataset* do repositório, os elementos metadados são classificados em: requeridos, fortemente recomendados, recomendados e opcionais.

*Elementos requeridos* são aqueles considerados ‘essenciais’ à descrição do *dataset*; são de preenchimento obrigatório e podem ser identificados visualmente por um asterisco (\*) após o nome de cada elemento (campo ou subcampo). Os elementos requeridos foram assim definidos pelo software Dataverse, e caso não sejam preenchidos o sistema impedirá a continuidade da catalogação do *dataset*. São os seguintes os elementos requeridos: Título; Autor (Nome); Contato (E-mail); Descrição (Texto); Assunto; e, Depositante.

*Elementos fortemente recomendados* não são de preenchimento obrigatório, visto que o software Dataverse não os considera como elementos requeridos. No entanto, quando preenchidos, os elementos fortemente recomendados fornecem informações adicionais indispensáveis à qualificação do *dataset*, contribuindo para a geração de metadados enriquecidos e aderentes aos princípios FAIR. Esses elementos e/ou campos são: Contato (Nome); Palavras-chave (Termo-chave); Palavras-chave (Vocabulário controlado do termo-chave); Palavras-chave (URL do termo-chave); Classificação (Termo de classificação); Classificação (Vocabulário controlado da classificação); Classificação (URL do termo de classificação); Data da produção; Colaborador (Nome); e, Distribuidor (Nome).

*Elementos recomendados* também não são de preenchimento obrigatório; no entanto, quando preenchidos, os elementos recomendados fornecem informações adicionais ‘relevantes’ à descrição do *dataset*, contribuindo igualmente para a geração de metadados com qualidade FAIR. Os elementos e/ou campos recomendados são: URL alternativa; Autor (Afiliação); Autor (Sistema identificador); Autor (Identificador); Contato (Afiliação); Publicação relacionada (Citação); Publicação relacionada (Tipo de identificador); Publicação relacionada (Número de identificador); Produtor (Nome do produtor); Produtor (Afiliação); Local de produção; Financiamento (Nome do órgão financiador); Financiamento (Número do contrato); Distribuidor (Afiliação); Data da distribuição; Data do depósito; Período de cobertura dos dados (Data de início); Período de cobertura dos dados (Data de término); Período de coleta dos dados (Data de início); Período de coleta dos dados (Data de término); Tipos de dados; Software (Nome); Software (Versão).

*Elementos opcionais* quando preenchidos fornecerão informações adicionais julgadas pertinentes à descrição do *dataset*, e de interesse do catalogador. Os elementos e/ou campos opcionais são: Título (Subtítulo); Título (Título alternativo); Outro identificador (Agência); Outro identificador (Identificador); Descrição (Data); Publicação relacionada (URL da publicação);

Notas; Idioma; Produtor (Abreviatura do nome do produtor); Produtor (URL do produtor); Produtor (URL do logotipo); Colaborador (Tipo); Distribuidor (Abreviatura); Distribuidor (URL do distribuidor); Distribuidor (URL do logotipo); Série (Nome); Série (Informação); Material relacionado; *Dataset* relacionado; Outras referências; Fontes de dados; Origem das fontes dos dados; Características das fontes observadas; e, Documentação e acesso a fontes de dados.

*Esclarecimento:* Em relação à exigência de preenchimento dos campos, a equipe do Projeto-piloto decidiu estabelecer mais dois níveis de exigência (fortemente recomendado e recomendado). O acréscimo desses dois níveis objetivou a ampliação da capacidade de recuperação, reúso e integração dos datasets com outros conjuntos e repositórios, de acordo com os princípios FAIR.

## Pontuação

Não usar ponto final no preenchimento dos campos, exceto para aqueles nos quais o fechamento do período é necessário, como é o caso dos elementos Descrição, Citação e Notas.


## Maiúsculas


Usar letra maiúscula inicial para o preenchimento de campos dos elementos Título e Palavras-chaves, e para nomes próprios. No elemento Descrição o uso de maiúscula deve obedecer às regras de redação. Nenhum elemento deve ser preenchido com letras maiúsculas, exceto siglas.

## Tecla *Enter*

A tecla '*Enter*' não deve ser usada.

## Botões de acesso

O botão  que aparece em alguns campos e/ou subcampos do formulário de catalogação indica que é possível adicionar mais de uma ocorrência durante a descrição de *datasets*.

O botão  indica que é possível excluir uma ocorrência de um campo e/ou subcampo; ou seja, com este botão é possível eliminar do formulário um campo ou subcampo preenchido anteriormente.

## Ajuda rápida

Para cada elemento metadado ou campo a ser preenchido dentro do formulário de catalogação está disponível uma “ajuda rápida”. Para acessar essa ajuda basta clicar sobre o ponto de interrogação posicionado à frente do elemento metadado ou campo.

## **Nomes pessoais e nomes corporativos, e-mails pessoais e e-mails corporativos e afiliação**

Para facilitar, agilizar e padronizar o preenchimento desses campos recomenda-se criar e manter, à parte, em arquivo texto ou planilha, relação de nomes pessoais e corporativos, de e-mails pessoais e corporativos e de afiliação para fins de registro e controle. Esses registros ajudarão na consistência e concisão das informações, evitando ambiguidades na descrição dos conteúdos do dataset.

### **Nome de autor pessoal**

Nome do autor quando este é uma pessoa. Deve ser descrito no formato de **sobrenome, nome**. Seguem exemplos:

Silva, José

Pereira Júnior, João

### **Nome de autor corporativo**

Nome do autor quando este é uma instituição. O preenchimento, neste caso, deve ser o nome da instituição, sem abreviação.

Exemplo:

Empresa Brasileira de Pesquisa Agropecuária

### **E-mail pessoal**

Endereço de e-mail de uma pessoa. Deve ser preenchido por um endereço válido, composto de duas partes: nome do usuário e nome do domínio, ligadas pelo caracter @  
Exemplo: usuario@provedor.com.br

### **E-mail corporativo**

Endereço de e-mail de uma instituição. Deve ser preenchido por um endereço válido, composto de duas partes: nome do usuário e nome do domínio, ligadas pelo caracter @  
Exemplo: usuario @provedor.com.br

### **Afiliação**

Nome de instituição onde está vinculada uma pessoa.

### **Palavras-chave – termos livres**

O uso de palavras-chave, de natureza termos livres, ou seja, termos que não constam dos vocabulários controlados adotados, requer registro e controle à parte, em arquivo texto ou planilha, de modo a organizar minimamente tais termos livres. Esse registro ajudará na consistência e no controle terminológico, a fim de evitar ambiguidades no tratamento do assunto do *dataset*. Também serão indicativos de termos a serem sugeridos aos vocabulários controlados.

## Data

Para o preenchimento de campos correspondentes a datas, adota-se a recomendação ISO 8601<sup>6</sup> (Date and Time Format), cujo formato é AAAA-MM-DD, onde AAAA é o ano, MM é o mês, e DD o dia.

## 3. Descrição dos Metadados de Citação

Para a catalogação de *datasets*, o software Dataverse possui um conjunto de elementos metadados, considerado *default*, e que são denominados de Metadados de Citação. Esse conjunto de metadados é formado por 33 elementos, dentre os quais incluem-se campos e subcampos.

Seguem listados os elementos:

- Título.
- URL alternativa.
- Outro identificador.
- Autor.
- Contato.
- Descrição.
- Assunto.
- Palavras-chave.
- Classificação.
- Publicação relacionada.
- Notas.
- Idioma.
- Produtor.
- Data de produção.
- Local de produção.
- Colaborador.
- Financiamento.
- Distribuidor.
- Data de distribuição.
- Depositante.
- Data de depósito.
- Período de cobertura de dados.
- Período de coleta de dados.
- Tipos de dados.

---

<sup>6</sup> Disponível em: <<https://www.w3.org/TR/NOTE-datetime>>



- Série.
- Software.
- Material selecionado.
- *Datasets* relacionados.
- Outras referências.
- Fontes de dados.
- Origem das fontes de dados.
- Características das fontes de dados observadas.
- Documentação e acesso a fontes de dados.

Para a catalogação de *datasets* foi adotado um conjunto de atributos comuns que orientam a descrição de todos os elementos metadados (campos e subcampos), conforme mostra a Tabela 1.

**Tabela 1.** Atributos orientadores para a descrição dos elementos (campos e subcampos) de Metadados de Citação

<b>Nome</b>	O rótulo atribuído para o elemento de dado.
<b>Identificador</b>	O único identificador atribuído para o elemento de dado.
<b>Ocorrência</b>	Indica qualquer limite para a repetibilidade do elemento de dado.
<b>Exigência de preenchimento</b>	Indica se o elemento de dado necessita estar presente sempre ou esporadicamente.
<b>Definição</b>	Uma declaração que claramente representa o conceito e a natureza essencial do elemento de dado.
<b>Comentário</b>	Um comentário relativo à aplicação do elemento de dado.
<b>Exemplo</b>	Um exemplo relativo à aplicação do elemento de dado.
<b>Documento orientador</b>	Um documento que oriente a aplicação do elemento de dado.

Fonte: Adaptação de Alves et al. (2008, p. 14).

A seguir são apresentados os 33 elementos de Metadados de Citação, incluindo campos e subcampos, com a respectiva descrição para a catalogação de *dataset*.

#### **Título\***

Identificador	Título
Ocorrência	Limitada
Exigência de preenchimento	Requerido
Definição	Nome atribuído ao <i>dataset</i> , e pelo qual será formalmente conhecido. Título deve ser formado pela estrutura a seguir: <b>Tipo de estudo ômico + Amostra estudada + Espécie estudada (nome comum + nome científico)</b> .

Comentário	Descrever o título do <i>dataset</i> observando as normas da língua culta, quanto à grafia e acentuação. Iniciais de título, de nome próprio e de nome científico devem ser grafadas em maiúscula.
Exemplo	Transcriptoma da glândula salivar do carrapato bovino ( <i>Rhipicephalus</i> (B). <i>microplus</i> ).
Documento orientador	RIBEIRO, A. M. de C. M. <b>Catálogo de recursos bibliográficos pela AACR2 2002</b> : Anglo-American Cataloguing Rules, 2nd edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular.

## Subtítulo

Identificador	Subtítulo
Ocorrência	Limitada
Exigência de preenchimento	Opcional
Definição	Nome secundário atribuído ao <i>dataset</i> para generalizar ou especificar a abrangência do título.
Comentário	Caso exista subtítulo, descreva-o de modo a ampliar ou restringir o título.
Exemplos	Título: Genoma do carrapato montado  Subtítulo: Expressão gênica na saliva do carrapato bovino
Documento orientador	RIBEIRO, A. M. de C. M. <b>Catálogo de recursos bibliográficos pela AACR2 2002</b> : Anglo-American Cataloguing Rules, 2nd edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular.

## Título alternativo

Identificador	Título alternativo
Ocorrência	Limitada
Exigência de preenchimento	Opcional
Definição	Nome alternativo para o <i>dataset</i> , sendo geralmente o nome pelo qual é comumente conhecido ou referido. Pode ser ainda uma forma abreviada do título.
Comentário	Caso exista, transcrevê-lo de modo a dar conhecimento à forma como o dataset também é conhecido.
Exemplo	Genoma do carrapato
Documento orientador	RIBEIRO, A. M. de C. M. <b>Catálogo de recursos bibliográficos pela AACR2 2002</b> : Anglo-American Cataloguing Rules, 2nd edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular.

## URL alternativa

Identificador	URL alternativa
Ocorrência	Recomendado
Exigência de preenchimento	Opcional

Definição	URL alternativa válida aponta para onde o <i>dataset</i> pode também ser visualizado. A URL não pode ser dependente de qualquer tipo de autenticação, uso de <i>captcha</i> , ou qualquer outro meio que impeça o acesso direto ao <i>dataset</i> .
Comentário	Transcrever a URL alternativa para acesso ao título do <i>dataset</i> .
Exemplos	<a href="https://www.frontiersin.org/articles/10.3389/fcimb.2019.00477/full#supplementary-material">https://www.frontiersin.org/articles/10.3389/fcimb.2019.00477/full#supplementary-material</a> <a href="https://www.ncbi.nlm.nih.gov/bioproject/PRJNA596777">https://www.ncbi.nlm.nih.gov/bioproject/PRJNA596777</a>
Documento orientador	Não se aplica.

## Outro identificador

Identificador	Outro identificador
Ocorrência	Ilimitada
Exigência de preenchimento	Opcional
Definição	Outro identificador único digital que identifica o <i>dataset</i> .
Comentário	É composto de dois subcampos: Agência e Identificador.
Exemplo	Não se aplica.
Documento orientador	Não se aplica.

## Agência

Identificador	Agência do identificador digital
Ocorrência	Ilimitada
Exigência de preenchimento	Opcional
Definição	Nome da agência que gerou o identificador digital.
Comentário	Transcrever o nome da agência que gerou o identificador digital.
Exemplo	NCBI
Documento orientador	Não se aplica.

## Identificador

Identificador	Identificador digital persistente
Ocorrência	Ilimitada
Exigência de preenchimento	Opcional
Definição	Outro Identificador digital que identifique o <i>dataset</i> .
Comentário	Transcrever outro identificador digital do <i>dataset</i> .
Exemplo	PRJNA596777

Documento orientador	Não se aplica.
----------------------	----------------

## Autor

Identificador	Autor
Ocorrência	Ilimitada
Exigência de preenchimento	Requerido
Definição	Nome da(s) pessoa(s) ou da instituição responsável(is) pela criação do <i>dataset</i> .
Comentário	É composto por quatro subcampos: Nome, Afiliação, Sistema identificador e Identificador.
Exemplo	Não se aplica.
Documento orientador	Não se aplica.

## Nome\*

Identificador	Nome
Ocorrência	Ilimitada
Exigência de preenchimento	Requerido
Definição	Nome(s) do(s) autor(es) do dataset. Pode(m) ser autor(es) pessoal(is) ou autor corporativo.
Comentário	Quando o autor é pessoal, preenche-se com o(s) nome(s) de pessoa(s), no formato sobrenome(s) (primeiro nome da família), seguidos dos prenomes. Caso seja um autor corporativo, descreve-se o nome da instituição, sem abreviações.
Exemplo	Nomes de autores pessoais: Silva, José Pereira Júnior, João  Nome de autor corporativo: Embrapa (nome síntese) Empresa Brasileira de Pesquisa Agropecuária
Documento orientador	RIBEIRO, A. M. de C. M. <b>Catálogo de recursos bibliográficos pela AACR2 2002</b> : Anglo-American Cataloguing Rules, 2nd edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular.

## Afiliação

Identificador	Afiliação
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Nome da instituição de vínculo do(s) autor(es) pessoal(is) durante o período de criação ou produção do <i>dataset</i> .
Comentário	Preencher com o nome completo da instituição. Não usar abreviações ou suprimir partes do nome. Caso o autor do dataset seja um autor corporativo, não é necessário preencher este campo.

Exemplo	Embrapa Gado de Corte Embrapa Informática Agropecuária
Documento orientador	RIBEIRO, A. M. de C. M. <b>Catálogo de recursos bibliográficos pela AACR2 2002</b> : Anglo-American Cataloguing Rules, 2nd edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular.

## Sistema do identificador

Identificador	Sistema do identificador digital do autor
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Nome do sistema do identificador digital escolhido para a identificação de cada autor pessoal do <i>dataset</i> . O uso de identificador digital persistente permite diferenciar um autor de outro. Também viabiliza a conexão automática entre o autor e suas atividades profissionais. Um sistema bastante utilizado é o ORCID ( <a href="https://orcid.org/">https://orcid.org/</a> ).
Comentário	Selecionar o sistema de identificação digital. Caso o autor seja corporativo, não deve ser preenchido.
Exemplo	ORCID ResearcherID Scopus ID
Documento orientador	Não se aplica.

## Identificador

Identificador	Identificador digital persistente
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Identificador digital persistente do(s) autor(es) pessoal(is) do <i>dataset</i> .
Comentário	A partir do sistema identificador selecionado, transcrever o identificador digital persistente do autor pessoal do <i>dataset</i> . Caso o autor seja corporativo, este campo não deve ser preenchido.
Exemplo	No esquema ORCID um exemplo é: <a href="http://orcid.org/0000-0002-1825-0097">http://orcid.org/0000-0002-1825-0097</a> (ORCID de Josiah Carberry)
Documento orientador	Não se aplica.

## Contato

Identificador	Contato
Ocorrência	Ilimitada
Exigência de preenchimento	Requerido
Definição	Nome da(s) pessoa(s) ou da instituição para contato sobre o <i>dataset</i> .
Comentário	É composto por três subcampos: Nome, Afiliação e E-mail.
Exemplo	Não se aplica.

Documento orientador	Não se aplica.
----------------------	----------------

## Nome

Identificador	Nome
Ocorrência	Ilimitada
Exigência de preenchimento	Fortemente recomendado
Definição	Nome pessoal ou corporativo para contato sobre o <i>dataset</i> .
Comentário	Nome pessoal deve ser descrito na forma direta, ou seja, prenome(s) seguidos de sobrenome(s). Caso o contato seja um nome corporativo, preencha com nome da instituição, sem abreviações.
Exemplo	Nome pessoal: José Silva João Pereira Júnior  Nome de corporativo para contato: Embrapa Informática Agropecuária
Documento orientador	RIBEIRO, A. M. de C. M. <b>Catálogo de recursos bibliográficos pela AACR2 2002</b> : Anglo-American Cataloguing Rules, 2nd edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular.

## Afiliação

Identificador	Afiliação do contato
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Nome da instituição à qual a pessoa indicada para contato está vinculada.
Comentário	Preencher com o nome completo da instituição. Não usar abreviações ou suprimir partes do nome. Caso a instituição já seja a indicada para contato sobre o <i>dataset</i> , não é necessário preencher este campo.
Exemplos	Embrapa Gado de Corte Embrapa Informática Agropecuária
Documento orientador	RIBEIRO, A. M. de C. M. <b>Catálogo de recursos bibliográficos pela AACR2 2002</b> : Anglo-American Cataloguing Rules, 2nd edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular.

## E-mail do contato\*

Identificador	E-mail
Ocorrência	Ilimitada
Exigência de preenchimento	Requerido
Definição	Endereço de e-mail do contato sobre o <i>dataset</i> , seja pessoal ou corporativo.
Comentário	Preencher com o e-mail do contato.

Exemplo	jose.silva@empresa.com.br
Documento orientador	Não se aplica.

## Descrição

Identificador	Descrição
Ocorrência	Ilimitada
Exigência de preenchimento	Requerido
Definição	Um sumário descritivo e sucinto contendo o propósito, a natureza e o escopo do <i>dataset</i> .
Comentário	É composto por dois subcampos: Texto e Data
Exemplo	Não se aplica
Documento orientador	Não se aplica.

## Texto\*

Identificador	Texto
Ocorrência	Ilimitada
Exigência de preenchimento	Requerido
Definição	Texto sucinto contendo o propósito, a natureza e o escopo do <i>dataset</i> .
Comentário	Texto livre.
Exemplo	Conjunto de dados réplica 1 da glândula salivar Holandês, dados brutos formato fastq.
Documento orientador	Não se aplica.

## Data

Identificador	Data
Ocorrência	Ilimitada
Exigência de preenchimento	Opcional
Definição	Data em que a descrição do <i>dataset</i> foi realizada. Um <i>dataset</i> pode ter mais de uma descrição, realizada em datas distintas.
Comentário	Seguir o padrão AAAA-MM-DD.
Exemplo	2020-06-08
Documento orientador	ISO 8601 Date and Time Format.

**Assunto\***

Identificador	Assunto
Ocorrência	Limitada
Exigência de preenchimento	Requerido
Definição	Áreas de domínio específicas e relevantes para o <i>dataset</i> .
Comentário	Selecionar a opção de área de domínio com as nas quais o <i>dataset</i> está relacionado. Pode-se selecionar mais de uma opção.
Exemplos	L Agricultural Sciences Medicine, Health and Life Sciences
Documento orientador	Não se aplica.

**Palavras-chave**

Identificador	Palavras-chave
Ocorrência	Ilimitada
Exigência de preenchimento	Fortemente recomendado
Definição	Palavras que expressam o conteúdo-chave do <i>dataset</i> , sendo capazes de representar aspectos relevantes para a compreensão do assunto. Palavras-chave podem ser utilizadas na construção de índices, na realização de buscas e na recuperação de informação. Palavras-chaves podem estar vinculadas ou não a um vocabulário controlado.
Comentário	É composto de três subcampos: Termo-chave, Vocabulário e URL do termo-chave.
Exemplo	Não se aplica.
Documentos orientadores	Não se aplica.

**Termo-chave**

Identificador	Termo-chave
Ocorrência	Ilimitada
Exigência de preenchimento	Fortemente recomendado
Definição	Termos-chave são indicados para representar aspectos importantes de um <i>dataset</i> . Podem ser usados na construção de índices, realização de busca e recuperação de informação. Recomenda-se a utilização de vocabulário controlado.
Comentário	Preencher com termos-chave que descrevem aspectos relevantes do <i>dataset</i> . Podem estar vinculados a um vocabulário controlado ou não.
Exemplos	Host-parasite relationships Nucleotide sequences <i>Rhipicephalus (Boophilus) microplus</i> Ticks Transcriptome Carrapato
Documentos orientadores	NAL Thesaurus



	NCBI Taxonomy AGROVOC Thesaurus
--	------------------------------------

## Vocabulário controlado do termo-chave

Identificador	Vocabulário controlado do termo-chave
Ocorrência	Ilimitada
Exigência de preenchimento	Fortemente recomendado
Definição	Vocabulário controlado utilizado para a atribuição de termos-chave do <i>dataset</i> .
Comentário	Preencher com nome do vocabulário controlado utilizado para a atribuição de termos-chave para o <i>dataset</i> .
Exemplos	NAL Thesaurus NCBI Taxonomy AGROVOC Thesaurus
Documento orientador	Não se aplica.

## URL do termo-chave

Identificador	URL do termo-chave
Ocorrência	Ilimitada
Exigência de preenchimento	Fortemente recomendado
Definição	URL do termo-chave onde o mesmo poderá ser acessado.
Comentário	Preencher com a URL válida do site onde se encontra a descrição do termo-chave.
Exemplos	URL do termo-chave 'nucleotide sequence' (AGROVOC Thesaurus): <a href="http://aims.fao.org/aos/agrovoc/c_27583">http://aims.fao.org/aos/agrovoc/c_27583</a>  URL do termo-chave 'nucleotide sequences' (NAL Thesaurus): <a href="https://agclass.nal.usda.gov/mtwdk.exe?k=default&amp;l=60&amp;w=15979&amp;s=5&amp;t=2">https://agclass.nal.usda.gov/mtwdk.exe?k=default&amp;l=60&amp;w=15979&amp;s=5&amp;t=2</a>
Documento orientador	Não se aplica.

## Classificação

Identificador	Classificação
Ocorrência	Ilimitada
Exigência de preenchimento	Fortemente recomendado
Definição	Refere-se à classificação do assunto do <i>dataset</i> , por meio da atribuição de uma ou mais categorias, sob as quais podem ser agrupados <i>datasets</i> de natureza semelhante. Recomenda-se o uso de categorias vinculadas a vocabulário controlado.
Comentário	É composto de três subcampos: Termo de classificação, Vocabulário controlado da classificação e URL do termo de classificação
Exemplo	Não se aplica.

Documento orientador	Não se aplica.
----------------------	----------------

### Termo de classificação

Identificador	Termo de classificação
Ocorrência	Ilimitada
Exigência de preenchimento	Fortemente recomendado
Definição	Indicação da classificação de assunto (categoria) de maior abrangência, na qual o <i>dataset</i> se classifica.
Comentário	Preencher com a expressão que indique a classificação de assunto mais abrangente para representar o <i>dataset</i> . O uso termos associados a vocabulários controlados é recomendado.
Exemplos	Animal Science and Animal Products (NAL Thesaurus) Phenomena (AGROVOC Thesaurus) Ixodida (NCBI Taxonomy)
Documento orientador	Não se aplica.

### Vocabulário controlado da classificação

Identificador	Vocabulário controlado da classificação
Ocorrência	Ilimitada
Exigência de preenchimento	Fortemente recomendado
Definição	Vocabulário controlado utilizado para a classificação de assunto do <i>dataset</i> .
Comentário	Preencher com nome do vocabulário controlado utilizado para a atribuição da classificação de assunto para representar o <i>dataset</i> .
Exemplos	NAL Thesaurus NCBI Taxonomy AGROVOC Thesaurus
Documento orientador	Não se aplica.

### URL do termo de classificação

Identificador	URL do termo de classificação
Ocorrência	Ilimitada
Exigência de preenchimento	Fortemente recomendado
Definição	URL do termo de classificação de assunto (categoria) na qual o <i>dataset</i> está classificado, de acordo com vocabulário controlado escolhido.
Comentário	Transcrever a URL do termo de classificação de assunto atribuída ao <i>dataset</i> .
Exemplos	URL do termo de classificação 'phenomena' (AGROVOC Thesaurus): <a href="http://aims.fao.org/aos/agrovoc/c_330704">http://aims.fao.org/aos/agrovoc/c_330704</a>

	<p>URL do termo de classificação 'L Animal Science and Animal Products' (NAL Thesaurus):  <a href="https://agclass.nal.usda.gov/mtwdk.exe?k=default&amp;l=60&amp;s=cid&amp;w=100">https://agclass.nal.usda.gov/mtwdk.exe?k=default&amp;l=60&amp;s=cid&amp;w=100</a></p> <p>URL do termo de classificação 'Ixodida' (NCBI Taxonomy):  <a href="https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&amp;id=6935&amp;lvl=3&amp;p=nucore&amp;lin=f&amp;keep=1&amp;srchmode=1&amp;unlock">https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&amp;id=6935&amp;lvl=3&amp;p=nucore&amp;lin=f&amp;keep=1&amp;srchmode=1&amp;unlock</a></p>
Documento orientador	Não se aplica.

## Publicação relacionada

Identificador	Publicação relacionada
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Refere-se a publicação que usa dados do dataset que está sendo catalogado.
Comentário	É composto de quatro subcampos: Citação, Tipo de identificador, Número do identificador e URL da publicação.
Exemplo	Não se aplica.
Documento orientador	Não se aplica.

## Citação

Identificador	Citação
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Refere-se à referência completa da publicação relacionada aos –dados do <i>dataset</i> que está sendo catalogado.
Comentário	Preencher em formato de referência bibliográfica a publicação relacionada ao <i>dataset</i> . Recomenda-se adotar o padrão ABNT 6023.
Exemplo	GIACHETTO, P. F. Transcriptoma do carrapato dos bovinos. In: GARCIA, M. V.; KOLLER, W. W. (Ed.). <b>Carrapatos na cadeia produtiva de bovinos</b> . Brasília, DF: Embrapa, 2019. cap. 16. p. 205-222.
Documento orientador	Manual de referência para recursos da informação da Embrapa. Disponível em: < <a href="https://www.embrapa.br/manual-de-referenciacao">https://www.embrapa.br/manual-de-referenciacao</a> >.

## Tipo de identificador

Identificador	Tipo de identificador
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Tipo de identificador digital usado pela publicação.

Comentário	Indicar a opção que corresponda ao tipo de identificador digital utilizado pela publicação relacionada ao <i>dataset</i> .
Exemplo	DOI
Documento orientador	Não se aplica

### Número do identificador

Identificador	Número do identificador
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Identificador digital único que remete para a publicação relacionada ao <i>dataset</i> .
Comentário	Transcrever o identificador digital único da publicação relacionada ao <i>dataset</i> .
Exemplo	Número do identificador digital: <a href="https://doi.org/10.3389/fcimb.2019.00477">https://doi.org/10.3389/fcimb.2019.00477</a>
Documento orientador	Não se aplica.

### URL da publicação

Identificador	URL da publicação
Ocorrência	Ilimitada
Exigência de preenchimento	Opcional
Definição	URL da página web onde se encontra a publicação. Pode ser a página de um periódico indexado, revista etc.
Comentário	Transcrever a URL da página web da publicação relacionada ao <i>dataset</i> .
Exemplo	URL da publicação: <a href="https://www.frontiersin.org/journals/cellular-and-infection-microbiology">https://www.frontiersin.org/journals/cellular-and-infection-microbiology</a>
Documento orientador	Não se aplica.

### Notas

Identificador	Notas
Ocorrência	Limitada
Exigência de preenchimento	Opcional
Definição	Informações adicionais sobre o <i>dataset</i> .
Comentário	Texto livre.
Exemplo	Réplica de dados de glândula salivar do carrapato bovino infestando gado Holandês. Ensaio envolve dados de gado Nelore e mestiço Nelore x Holandês.
Documento orientador	Não se aplica.

## Idioma

Identificador	Idioma
Ocorrência	Ilimitada
Exigência de preenchimento	Opcional
Definição	Idioma do <i>dataset</i> .
Comentário	Selecionar um ou mais idiomas a partir da lista de opções oferecidas.
Exemplos	Inglês Inglês Português Inglês Português Espanhol
Documento orientador	Não se aplica.

## Produtor

Identificador	Produtor
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Nome pessoal ou nome corporativo do responsável financeiro ou administrativo pelo <i>dataset</i> .
Comentário	É composto por cinco subcampos: Nome do produtor, Afiliação do produtor, Abreviatura do nome do produtor, URL do produtor e URL do logotipo do produtor.
Exemplo	Não se aplica.
Documento orientador	Não se aplica.

## Nome do produtor

Identificador	Nome do produtor
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Nome pessoal ou corporativo do produtor do <i>dataset</i> .
Comentário	Nome pessoal deve ser descrito na forma direta, ou seja, prenomes seguidos de sobrenomes. Caso o produtor seja um nome corporativo, preencher com nome da instituição, sem abreviações.
Exemplos	Nome pessoal: José Silva  Nome de corporativo Empresa Brasileira de Pesquisa Agropecuária
Documento orientador	RIBEIRO, A. M. de C. M. <b>Catalogação de recursos bibliográficos pela AACR2 2002</b> : Anglo-American Cataloguing Rules, 2nd edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular.

### Afiliação do produtor

Identificador	Afiliação do produtor
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Nome da instituição à qual o produtor do <i>dataset</i> está vinculado.
Comentário	Preencher com o nome completo da instituição. Não usar abreviações ou suprimir partes do nome. Caso a instituição já seja a indicada como o produtor do <i>dataset</i> , não é necessário preencher este campo.
Exemplos	Embrapa Gado de Corte Empresa Brasileira de Pesquisa Agropecuária
Documento orientador	RIBEIRO, A. M. de C. M. <b>Catálogo de recursos bibliográficos pela AACR2 2002</b> : Anglo-American Cataloguing Rules, 2nd edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular.

### Abreviatura do nome do produtor

Identificador	Abreviatura do nome do produtor
Ocorrência	Ilimitada
Exigência de preenchimento	Opcional
Definição	Abreviatura do nome do produtor do <i>dataset</i> .
Comentário	Indicar a abreviatura do nome do produtor quando se tratar de um nome corporativo.
Exemplos	Embrapa LMB
Documento orientador	Não se aplica.

### URL do produtor

Identificador	URL do produtor
Ocorrência	Ilimitada
Exigência de preenchimento	Opcional
Definição	URL válida de uma página na web do produtor do <i>dataset</i> .
Comentário	Transcrever a URL válida da página web do produtor do <i>dataset</i> .
Exemplos	URL Empresa Brasileira de Pesquisa Agropecuária: <a href="https://www.embrapa.br/">https://www.embrapa.br/</a>  URL do Laboratório Multiusuário de Bioinformática: <a href="https://www.embrapa.br/laboratorio-de-bioinformatica">https://www.embrapa.br/laboratorio-de-bioinformatica</a>
Documento orientador	Não se aplica.

## URL do logotipo do produtor

Identificador	URL do logotipo do produtor
Ocorrência	Ilimitada
Exigência de preenchimento	Opcional
Comentário	Transcrever uma URL válida da página web onde o logotipo do produtor pode ser visualizado.
Exemplos	<p>URL do logotipo da Embrapa:  <a href="https://www.embrapa.br/image/layout_set_logo?img_id=1343656&amp;t=1591812010844">https://www.embrapa.br/image/layout_set_logo?img_id=1343656&amp;t=1591812010844</a></p> <p>URL do logotipo do Laboratório Multiusuário de Bioinformática:  <a href="https://www.dataverse-h.cnptia.embrapa.br/logos/8/layout_set_logo.png">https://www.dataverse-h.cnptia.embrapa.br/logos/8/layout_set_logo.png</a></p>
Documento orientador	Não se aplica.

## Data de produção

Identificador	Data de produção
Ocorrência	Limitada
Exigência de preenchimento	Fortemente recomendado
Definição	Data em que o <i>dataset</i> foi produzido.
Comentário	Seguir o padrão AAAA-MM-DD.
Exemplo	2015-05-01
Documento orientador	ISO 8601 Date and Time Format.

## Local de produção

Identificador	Local de produção
Ocorrência	Limitada
Exigência de preenchimento	Recomendado
Definição	Local onde o <i>dataset</i> e quaisquer outros materiais relacionados foram produzidos.
Comentário	Texto livre.
Exemplo	Laboratório de Biotecnologia Animal Centro de Genômica Funcional
Documento orientador	Não se aplica.

## Colaborador

Identificador	Colaborador
Ocorrência	Ilimitada
Exigência de preenchimento	Fortemente recomendado

Definição	Pessoa ou instituição que colabora com a coleta, gestão ou qualquer outra contribuição relevante para o desenvolvimento do <i>dataset</i> .
Comentário	É composto por dois subcampos: Tipo de colaborador e Nome do colaborador.
Exemplo	Não se aplica.
Documento orientador	Não se aplica.

### Tipo de colaborador

Identificador	Tipo de colaborador
Ocorrência	Ilimitada
Exigência de preenchimento	Opcional
Definição	Tipo de colaborador do <i>dataset</i> .
Comentário	Selecionar um tipo de colaborador a partir da lista de opções oferecidas. Pode-se selecionar mais de um tipo de colaborador, mas para isso, é necessário adicionar mais uma ocorrência.
Exemplos	Coletor de dados Curador de dados Líder do projeto Pesquisador
Documento orientador	Não se aplica.

### Nome do colaborador

Identificador	Nome do colaborador
Ocorrência	Ilimitada
Exigência de preenchimento	Fortemente recomendado
Definição	Nome pessoal ou corporativo do colaborador do <i>dataset</i> .
Comentário	Nomes pessoais devem ser indicados na forma direta, ou seja, prenomes seguidos de sobrenomes. Caso o nome do colaborador seja um nome corporativo, preencher o nome completo da instituição; não usar abreviaturas ou suprimir de partes do nome da instituição.
Exemplos	Nome pessoal de colaborador: Paula Regina Kuser Falcão  Nome de corporativo (instituição) de colaborador: Laboratório de Biotecnologia Animal Centro de Genômica Funcional
Documento orientador	Não se aplica.

### Financiamento

Identificador	Financiamento
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado



Definição	Informações sobre instituição, agência de fomento, organização e/ou projeto responsável pela concessão de suporte e/ou apoio financeiro para a produção.
Comentário	É composto pelos subcampos: Nome do órgão financiador e Número do contrato
Exemplo	Não se aplica.
Documento orientador	Não se aplica.

### Nome do órgão financiador

Identificador	Nome do órgão financiador
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Nome da instituição, agência de fomento ou organização que concedeu suporte e/ou apoio financeiro para a produção do <i>dataset</i> .
Comentário	Texto livre.
Exemplos	Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) Empresa Brasileira de Pesquisa Agropecuária (Embrapa) Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp)
Documento orientador	Não se aplica.

### Número do contrato

Identificador	Número do contrato
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Número do contrato de concessão de financiamento dos recursos para a produção do <i>dataset</i> .
Comentário	Transcrever o número do contrato, caso exista.
Exemplo	01.11.07.002.03.00 (projeto SEG, da Embrapa)
Documento orientador	Não se aplica.

### Distribuidor

Identificador	Distribuidor
Ocorrência	Ilimitada
Exigência de preenchimento	Fortemente recomendado
Definição	Pessoa ou instituição designada pelo autor ou produtor do <i>dataset</i> para gerar cópias do trabalho, incluindo quaisquer edições ou revisões necessárias.
Comentário	É composto de cinco subcampos: Nome do distribuidor, Afiliação do distribuidor, Abreviatura do nome distribuidor, URL do distribuidor e URL do logotipo do distribuidor

Exemplo	Não se aplica.
Documento orientador	Não se aplica.

### Nome do distribuidor

Identificador	Nome do distribuidor
Ocorrência	Ilimitada
Exigência de preenchimento	Fortemente recomendado
Definição	Nome pessoal ou corporativo do distribuidor do <i>dataset</i> .
Comentário	Nome pessoal deve ser descrito na forma direta, ou seja, prenome(s) seguido(s) de sobrenome(s). Caso o distribuidor seja um nome corporativo, preencha com nome da instituição, sem abreviações.
Exemplos	Nome pessoal: José Silva  Nome corporativo Empresa Brasileira de Pesquisa Agropecuária
Documento orientador	Não se aplica.

### Afiliação do distribuidor

Identificador	Afiliação do distribuidor
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Nome da instituição à qual o distribuidor do <i>dataset</i> está vinculado.
Comentário	Preencher com o nome completo da instituição. Não usar abreviações ou suprimir partes do nome. Caso a instituição já seja a indicada como o distribuidor do <i>dataset</i> , não é necessário preencher este campo.
Exemplos	Embrapa (nome síntese) Empresa Brasileira de Pesquisa Agropecuária
Documento orientador	Não se aplica.

### Abreviatura do nome do distribuidor

Identificador	Abreviatura do nome do distribuidor
Ocorrência	Ilimitada
Exigência de preenchimento	Opcional
Definição	Abreviatura do nome distribuidor do <i>dataset</i> .
Comentário	Indicar a abreviatura do nome do distribuidor quando se tratar de um nome corporativo.

Exemplos	ESALQ LMB
Documento orientador	Não se aplica.

### URL do distribuidor

Identificador	URL do distribuidor
Ocorrência	Ilimitada
Exigência de preenchimento	Opcional
Definição	URL válida de uma página na web do distribuidor do <i>dataset</i> .
Comentário	Transcrever a URL válida da página web do distribuidor do <i>dataset</i> .
Exemplos	URL da Empresa Brasileira de Pesquisa Agropecuária: <a href="https://www.embrapa.br/">https://www.embrapa.br/</a>  URL do Laboratório Multiusuário de Bioinformática: <a href="https://www.embrapa.br/laboratorio-de-bioinformatica">https://www.embrapa.br/laboratorio-de-bioinformatica</a>
Documento orientador	Não se aplica.

### URL do logotipo do distribuidor

Identificador	URL do logotipo do distribuidor
Ocorrência	Ilimitada
Exigência de preenchimento	Opcional
Definição	URL válida do logotipo do distribuidor do <i>dataset</i> .
Comentário	Transcrever a URL válida da página web onde o logotipo do distribuidor pode ser visualizado.
Exemplos	URL do logotipo da Embrapa: <a href="https://www.embrapa.br/image/layout_set_logo?img_id=1343656&amp;t=1591812010844">https://www.embrapa.br/image/layout_set_logo?img_id=1343656&amp;t=1591812010844</a>  URL do logotipo do Laboratório Multiusuário de Bioinformática: <a href="https://www.dataverse-h.cnptia.embrapa.br/logos/8/layout_set_logo.png">https://www.dataverse-h.cnptia.embrapa.br/logos/8/layout_set_logo.png</a>
Documento orientador	Não se aplica.

### Data de distribuição

Identificador	Data de distribuição
Ocorrência	Limitada
Exigência de preenchimento	Recomendado
Definição	Data em que o <i>dataset</i> foi disponibilizado para distribuição.
Comentário	Seguir o padrão AAAA-MM-DD.
Exemplo	2020-01-15
Documento orientador	ISO 8601 Date and Time Format

**Depositante\***

Identificador	Depositante
Ocorrência	Limitada
Exigência de preenchimento	Requerido Este campo é preenchido automaticamente pelo sistema, com o nome do autor do <i>dataset</i> .
Definição	Nome pessoal ou corporativo do depositante do <i>dataset</i> .
Comentário	Nome pessoal deve ser descrito na forma direta, ou seja, prenome(s) seguido(s) de sobrenome(s). Caso o depositante seja um nome corporativo, preencha com nome da instituição, sem abreviações.
Exemplos	Nome pessoal: José Silva  Nome corporativo: Empresa Brasileira de Pesquisa Agropecuária
Documento orientador	Não se aplica.

**Data do depósito**

Identificador	Data do depósito
Ocorrência	Limitada
Exigência de preenchimento	Recomendado
Definição	Data em que o <i>dataset</i> foi depositado no repositório.
Comentário	Seguir o padrão AAAA-MM-DD.
Exemplo	2020-06-11
Documento orientador	ISO 8601 Date and Time Format.

**Período da cobertura de dados**

Identificador	Período da cobertura de dados
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Período de tempo ao qual os dados se referem ou período de tempo abrangido pelos dados.
Comentário	É composto por dois subcampos: Data de início e Data de término
Exemplo	Não se aplica.
Documento orientador	Não se aplica.

**Data de início**

Identificador	Data de início
---------------	----------------

Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Data de início do período de cobertura dos dados.
Exemplo	2015-05-01
Documento orientador	ISO 8601 Date and Time Format.

## Data de término

Identificador	Data de término
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Data de término do período da cobertura dos dados.
Comentário	Seguir o padrão AAAA-MM-DD.
Exemplo	2015-06-22
Documento orientador	ISO 8601 Date and Time Format.

## Período da coleta de dados

Identificador	Período da coleta de dados
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Período de tempo em que os dados foram coletados.
Comentário	É composto por dois subcampos: Data de início e Data de término
Exemplo	Não se aplica.
Documento orientador	Não se aplica.

## Data de início

Identificador	Data de início
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Data de início do período da coleta dos dados.
Comentário	Seguir o padrão AAAA-MM-DD.
Exemplo	2015-05-01
Documento orientador	ISO 8601 Date and Time Format.

## Data de término

Identificador	Data de término
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Data de término do período da coleta de dados.
Comentário	Seguir o padrão AAAA-MM-DD.
Exemplo	2015-06-22
Documento orientador	ISO 8601 Date and Time Format.

## Tipo de dados

Identificador	Tipos de dados
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Tipos de dados do <i>dataset</i> , que podem ser identificados como: <i>survey</i> , dados de censo, dados agregados, dados clínicos, dados de eventos, dados de transações, código-fonte do programa, texto legível por máquina, dados de transações administrativas, dados experimentais, dados de teste psicológico, dados textuais, documentos codificados, dados/avaliações de observação, dados produzidos pelo processo etc.
Comentário	Texto livre. Indique textualmente o tipo de dado que corresponde ao <i>dataset</i> . Pode ser usado mais de um tipo.
Exemplos	Dados de pesquisa Dados experimentais Dados biológicos Dados genômicos Dados de sequência RNA
Documento orientador	Não se aplica.

## Série

Identificador	Série
Ocorrência	Limitada
Exigência de preenchimento	Opcional
Definição	Informação sobre a série do <i>dataset</i> , caso exista.
Comentário	Texto livre. É composto de dois subcampos: Nome da série e Informação sobre a série.
Exemplo	Não se aplica.
Documento orientador	Não se aplica.

## Nome da série

Identificador	Nome da série
Ocorrência	Limitada
Exigência de preenchimento	Opcional
Definição	Nome de série ao qual o <i>dataset</i> está vinculado.
Comentário	Texto livre.
Exemplo	Portfólio de Sanidade Animal e Manejo de Pragas e Doenças
Documento orientador	Não se aplica.

## Informação sobre série

Identificador	Informação sobre a série
Ocorrência	Limitada
Exigência de preenchimento	Opcional
Definição	Informação sucinta sobre a série à qual o <i>dataset</i> pertence.
Comentário	Texto livre.
Exemplo	A Série Portfólio Sanidade Animal e Manejo de Pragas e Doenças abarca o conjunto de <i>dataset</i> produzidos no âmbito do Projeto RmVac.
Documento orientador	Não se aplica.

## Software

Identificador	Software
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Informações sobre o software usado para gerar o <i>dataset</i> .
Comentário	Texto livre. É composto de dois subcampos: Nome do software e Versão do software.
Exemplo	Não se aplica.
Documento orientador	Não se aplica.

## Nome do software

Identificador	Nome do software
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Nome do software utilizado para gerar o <i>dataset</i> .
Comentário	Texto livre.

Exemplos	SAS <i>Scripts</i> Software R Trinity
Documento orientador	Não se aplica.

### Versão do software

Identificador	Versão
Ocorrência	Ilimitada
Exigência de preenchimento	Recomendado
Definição	Versão do software utilizado para gerar o <i>dataset</i> .
Comentário	Texto livre.
Exemplos	Versão 2.8.4 Versão 0.01
Documento orientador	Não se aplica.

### Material relacionado

Identificador	Material relacionado
Ocorrência	Ilimitada
Exigência de preenchimento	Opcional
Definição	Outro material relacionado ao dataset que está sendo descrito.
Comentário	Texto livre.
Exemplos	Projetos anteriores <i>Datasets</i> outros
Documento orientador	Não se aplica

### *Datasets* relacionados

Identificador	<i>Datasets</i> relacionados
Ocorrência	Ilimitada
Exigência de preenchimento	Opcional
Definição	Quaisquer <i>datasets</i> relacionados ao <i>dataset</i> que está sendo descrito, tais como <i>datasets</i> provenientes de pesquisas anteriores sobre esse assunto
Comentário	Texto livre.
Exemplo	Projetos anteriores
Documento orientador	Não se aplica.



## Outras referências

Identificador	Outras referências
Ocorrência	Ilimitada
Exigência de preenchimento	Opcional
Definição	Quaisquer outras referências que podem ter servido de material de apoio ao <i>dataset</i> .
Comentário	Texto livre.
Exemplo	Acesso e descoberta de conhecimento produzido por pesquisadores de instituições parceiras do Projeto RmVac.
Documento orientador	Não se aplica.

## Fontes de dados

Identificador	Fontes de dados
Ocorrência	Ilimitada
Exigência de preenchimento	Opcional
Definição	Referem-se a materiais que serviram de fontes para o <i>datasets</i> , como: lista de livros, artigos, publicações em série ou arquivos de dados legíveis por máquina.
Comentário	Texto livre.
Exemplo	GIACHETTO, P. F.; CUNHA, R. C.; NHANI JUNIOR, A.; GARCIA, M. V.; FERRO, J. A.; ANDREOTTI, R. Gene expression in the salivary gland of <i>Rhipicephalus (Boophilus) microplus</i> fed on tick-susceptible and tick-resistant hosts. <b>Frontiers in Cellular and Infection Microbiology</b> , v. 9, p. 1-15, Jan. 2020. DOI: 10.3389/fcimb.2019.00477.
Documento orientador	Manual de referência para recursos da informação da Embrapa. Disponível em: < <a href="https://www.embrapa.br/manual-de-referenciacao">https://www.embrapa.br/manual-de-referenciacao</a> >

## Origem das fontes de dados

Identificador	Origem das fontes de dados
Ocorrência	Limitada
Exigência de preenchimento	Opcional
Definição	Para materiais históricos, informar a origem das fontes e as regras seguidas no estabelecimento das fontes do <i>dataset</i> .
Comentário	Texto livre.
Exemplo	Literatura que retrata o estado da arte.
Documento orientador	Não se aplica

### Características das fontes de dados observadas

Identificador	Características das fontes de dados
Ocorrência	Limitada
Exigência de preenchimento	Opcional
Definição	Avaliação das características e material de origem do <i>dataset</i> . Descrever aspecto digno de nota para o dado coletado.
Comentário	Texto livre.
Exemplo	Não há
Documento orientador	Não se aplica.

### Documentação e acesso a fontes de dados

Identificador	Documentação e acesso a fontes de dados
Ocorrência	Limitada
Exigência de preenchimento	Opcional
Definição	Nível de acesso à documentação do material de origem. Informar quaisquer restrições de acesso aos dados de origem.
Comentário	Texto livre.
Exemplo	Não há.
Documento orientador	Não se aplica.

## 4. Descrição de Metadados *Life Science*

O software Dataverse também possibilita a catalogação de *dataset* no conjunto de elementos metadados denominado *Life Science*. Esse conjunto de metadados é formado por nove elementos, listados a seguir:

- Tipo de design.
- Tipo de fator.
- Organismo.
- Outro organismo.
- Tipo de medida.
- Outro tipo de medida.
- Tipo de tecnologia.
- Plataforma tecnológica.
- Tipo de célula.

Para a catalogação de *datasets* de acordo com os elementos Metadados *Life Science*, foi adotado um conjunto de atributos comuns que orientam a descrição de todos os seus elementos (campos e subcampos), conforme mostra a Tabela 2, abaixo.

**Tabela 2.** Atributos orientadores para a descrição dos elementos Metadados *Life Science*.

<b>Nome</b>	O rótulo atribuído para o elemento de dado.
<b>Identificador</b>	O único identificador atribuído para o elemento de dado.
<b>Ocorrência</b>	Indica qualquer limite para a repetibilidade do elemento de dado.
<b>Exigência de preenchimento</b>	Indica se o elemento de dado necessita estar presente sempre ou esporadicamente.
<b>Definição</b>	Uma declaração que claramente representa o conceito e a natureza essencial do elemento de dado.
<b>Comentário</b>	Um comentário relativo à aplicação do elemento de dado.
<b>Exemplo</b>	Um exemplo relativo à aplicação do elemento de dado.
<b>Documento orientador</b>	Um documento que oriente a aplicação do elemento de dado.

Fonte: Adaptação de Alves et al. (2008, p. 14).

A seguir são apresentados os nove elementos Metadados *Life Science* e respectiva descrição para a catalogação de *dataset*.

### Tipo de design

Identificador	Tipo de design
Ocorrência	Limitada
Exigência de preenchimento	Opcional
Definição	Definição geral do desenho experimental.
Comentário	Selecionar um ou mais tipos de design.
Exemplos	Case control; Cross sectional; Cohort study; Nested case control design; Not specified; Parallel group design; Perturbation design; Randomized controlled trial; Technological design.
Documento orientador	Não se aplica.

## Tipo de fator

Identificador	Tipo de fator
Ocorrência	Limitada
Exigência de preenchimento	Opcional
Definição	Fatores utilizados no <i>dataset</i> . Fator experimental; estresse hídrico; estágio de desenvolvimento; tipo de célula.
Comentário	Selecionar um ou mais tipos de fator.
Exemplos	Age; Biomarkers; Cell surface markers; Cell type/cell line; Developmental stage; Genetic characteristics; Treatment type etc.
Documento orientador	Não se aplica.

## Organismo

Identificador	Organismo
Ocorrência	Limitada
Exigência de preenchimento	Opcional
Definição	Nome taxonômico do organismo usado nos <i>datasets</i> ou de onde o material biológico se originou.
Comentário	Selecionar um ou mais organismos. Para indicar organismos não constantes na lista apresentada, selecione 'Other'.
Exemplos	<i>Bos taurus</i> <i>Escherichia coli</i> <i>Rhipicephalus microplus</i> Carrapato do boi
Documento orientador	Não se aplica.

## Outro organismo

- Não se aplica

## Tipo de medida

Identificador	Tipo de medida
Ocorrência	Limitada
Exigência de preenchimento	Opcional
Definição	Um termo para qualificar o que está sendo medido.
Comentário	Selecionar um ou mais tipos de medida. Para indicar tipo de medida não constante na lista apresentada, selecione 'Other'.
Exemplos	Cell counting; DNA methylation profiling; Genome sequencing; Protein=RNA binding (RIP-Seq) etc.
Documento orientador	Não se aplica.

## Outro tipo de medida

- Não se aplica

## Tipo de tecnologia

Identificador	Tipo de tecnologia
Ocorrência	Limitada
Exigência de preenchimento	Opcional
Definição	Um termo usado para identificar a tecnologia usada para coletar os dados.
Comentário	Selecionar um ou mais tipos de tecnologia. Para indicar tipo de tecnologia não constante na lista apresentada, selecione 'Other'.
Exemplos	DNA microarray; Mass spectrometry; Nucleotide sequencing; Protein microarray etc.
Documento orientador	Não se aplica.

## Plataforma tecnológica

Identificador	Plataforma tecnológica
Ocorrência	Limitada
Exigência de preenchimento	Opcional
Definição	O fabricante e o nome da plataforma usada no experimento para coleta de dados.
Comentário	Selecionar um ou mais plataformas tecnológicas. Para indicar plataforma tecnológica não constante na lista apresentada, selecione 'Other'.
Exemplos	Illumina HiSeq2500
Documento orientador	Não se aplica.

## Tipo de célula

Identificador	Linhagem de células
Ocorrência	Ilimitada
Exigência de preenchimento	Opcional
Definição	Nome da linhagem de células das quais as amostras derivam.
Comentário	Texto livre. Indicar os nomes dos tipos de célula presentes nos tecidos, por exemplo, do carrapato.
Exemplos	Larvas; Ninfas; Ovários; Intestinos; Glândula salivar etc.
Documento orientador	Não se aplica.

## Referências

- ALVES, M. das D. R.; SOUZA, M. I. F.; SANTOS, A. D. dos. **Manual de representação descritiva de recursos eletrônicos no Sistema Agência**. Campinas: Embrapa Informática Agropecuária, 2008. 139 p. (Embrapa Informática Agropecuária. Documentos, 81). Disponível em: <<https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/48773>>. Acesso em: 17 set. 2020.
- CESAR JUNIOR, R. M. Apresentação à edição brasileira. In: HEY, T.; TANSLEY, S.; TOLLE, K. (Org.). **O quarto paradigma: descobertas científicas na era da eScience**. São Paulo: Oficina de Textos, 2011. p. 7.
- CHAN, L.; OKUNE, A.; SAMBULI, N. O que é ciência aberta e colaborativa, e que papéis ela poderia desempenhar no desenvolvimento? In: ALBABLI, S.; MACIEL, M. L.; ABDO, A. H. (Org.). **Ciência aberta, questões abertas**. Brasília, DF: Ibict; Rio de Janeiro: Unirio, 2015. cap. 5, p. 91-119. Disponível em: <<http://ridi.ibict.br/handle/123456789/465>>. Acesso em: 28 ago. 2020.
- CÓDIGO de catalogação anglo-americano. 2. ed., rev. 2002. São Paulo: FEBAB: Imprensa Oficial do Estado de São Paulo, 2004. 1 v. Paginação irregular
- CURTY, R. Abordagens de reuso e a questão de reusabilidade dos dados científicos. **Liinc em Revista**, v. 15, n. 2, p. 177-193, nov. 2019. DOI: <https://doi.org/10.18617/liinc.v15i2.4777>
- DATAVERSE. **Harvard Dataverse**. [Cambridge, MA: Harvard College, 2020]. Disponível em: <<https://dataverse.harvard.edu/>>. Acesso em: 2 maio 2020.
- EMBRAPA. **Manual de referência para recursos da informação da Embrapa**. 5. ed. Brasília, DF, 2020. Disponível em: <<https://www.embrapa.br/manual-de-referenciacao>>. Acesso em: 18 out. 2020.
- GO FAIR. **FAIR principles**. [2016]. Disponível em: <<https://www.go-fair.org/fair-principles/>>. Acesso em: 27 ago. 2020.
- GORDANA, R.; DRAGAN, I. Cataloguing dataset in library information systems using the MARC 21 format. In: INTERNATIONAL CONFERENCE ON INFORMATION SOCIETY AND TECHNOLOGY, 7., 2017, Kopaonik. **Proceedings...** Belgrade: Society for Information Systems and Computer Networks Belgrade, 2017. p. 395-399. Disponível em: <<https://pdfs.semanticscholar.org/407a/9d09ecb1261db980ef03b8025fe5dc1939d5.pdf>>. Acesso em: 28 ago. 2020.
- GRAY, J. Jim Gray on eScience: a transformed scientific method. In: HEY, T.; TANSLEY, S.; TOLLE, K. (Ed.). **The fourth paradigm: data-intensive scientific discovery**. Redmond: Microsoft Research, 2009. p. xvii-xxxi.
- HENNING, P. C.; RIBEIRO, C. J. S.; SALES, L. F.; MOREIRA, J. L. R.; SANTOS, L. O. B. da S. Desmistificando os princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, v. 14, n. 3, p. 175-192, 2019. DOI: 10.22478/ufpb.1981-0695.2019v14n3.46969.

RIBEIRO, A. M. de C. M. **AACR2**: Anglo-American cataloguing rules, 2nd edition: descrição e pontos de acesso. Brasília, DF: Ed. do Autor, 1995. 577 p.

RIBEIRO, A. M. de C. M. **Catálogo de recursos bibliográficos pela AACR2 2002**: Anglo-American Cataloguing Rules, 2nd edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa**: [minicurso]. [Rio de Janeiro: CNEN, 2016]. 196 slides.

SEMELER, A. R.; PINTO, A. L. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados. **Ciência da Informação**, v. 48, n. 1, p. 113-129, jan./abr. 2019.

VEIGA, V. Gestão de dados de pesquisa FAIR: dando um JUMP em seus dados. In: ENCONTRO DA REDE SUDESTE DE REPOSITÓRIOS INSTITUCIONAIS, 1., 2019, Rio de Janeiro. **Anais...** Rio de Janeiro: UFRJ, 2019. 59 p. Disponível em: <<https://www.arca.fiocruz.br/handle/icit/33343>>. Acesso em: 8 jul. 2020.

## Apêndice A - Vocabulários Controlados

### Introdução

Este Apêndice A traz orientações para o uso de vocabulários controlados no Repositório de Dados da Embrapa (AgDados) a partir da experiência do Projeto-piloto de Implantação de Gestão de Dados de Pesquisa no Laboratório Multiusuário de Bioinformática da Embrapa (LMB).

Vocabulário controlado consiste de uma lista de termos elaborada por especialistas de diversos domínios para indexação de assunto de diferentes objetos de informação, dentre os quais, os *datasets*. Vocabulário controlado é comumente definido como uma lista autorizada de assunto, um vocabulário fechado, um tesauro. Cunha e Cavalcanti (2008) assim definem vocabulários controlados:

Conjunto de termos que, nos sistemas de informação, devem ser empregados tanto no momento da indexação como no da recuperação. A finalidade principal desse controle é fazer coincidir a linguagem do pesquisador com a do indexador. Nos vocabulários controlados são feitas remissivas dos sinônimos e quase-sinônimos para o termo selecionado como descritor. (Cunha; Cavalcanti, 2008, p. 378).

Um dos objetivos de um vocabulário controlado é compatibilizar a linguagem utilizada por pesquisadores à linguagem adotada em sistemas de informação. Em processos de catalogação, os vocabulários controlados são utilizados para representar, por meio de termos ou palavras-chave, o conteúdo do recurso de informação que está sendo catalogado.

A adoção de vocabulário para descrever e classificar assuntos é uma medida necessária para prover serviços de busca e recuperação de dados e informações com maior garantia de qualidade. No contexto deste trabalho, o uso de vocabulário controlado tem a função de:

- Atribuir palavras-chave e classificação de assuntos na catalogação de *datasets*.
- Indicar relacionamentos e associações hierárquicas e de equivalência entre termos atribuídos para palavras-chave e classificação, atribuídos pelo catalogador de *datasets*.
- Eliminar polissemia e ambiguidade na atribuição termos nos campos Palavras-chave e Classificação.
- Possibilitar a identificação e recuperação de *datasets* escritos sob determinados termos.
- Explorar a riqueza terminológica existente nos diferentes vocabulários, sob a perspectiva da análise semântica e de mineração de textos.
- Oferecer possibilidade de análise conceitual dos termos em diferentes idiomas, subsidiando a descrição de assunto e de classificação de *datasets*.

Vocabulários controlados são altamente recomendados para a representação temática de *datasets*, sobretudo quando são compatíveis com os princípios FAIR, especial-



mente com o princípio *Interoperable* (Interoperabilidade) (Henning et al., 2019; Veiga, 2019). Metadados bem descritos com vocabulários controlados e indexados em catálogos e repositórios de dados são facilmente ‘encontráveis’ por mecanismos de busca. Quando isso acontece, tanto dados como metadados estarão ao alcance dos usuários. Na mesma direção, metadados descritos com vocabulários controlados permitem a representação de assuntos e de classificação de forma padronizada, garantindo, assim, a acessibilidade e a interoperabilidade semântica e legível por máquina (Veiga, 2019).

Nos campos e subcampos utilizados no software Dataverse, é possível o uso de termos de vocabulário controlado em Palavras-chave e também em Classificação. Em Palavras-chave, recomenda-se usar o termo mais específico possível para classificar o dataset. Em Classificação, recomenda-se utilizar um termo mais abrangente.

Os principais vocabulários controlados estão disponíveis na internet com acesso gratuito e consulta e uso on-line. No entanto, o uso de vocabulário controlado para identificar um termo ou classificação de assunto quase sempre exige o auxílio de um profissional da área de Biblioteconomia ou Ciência da Informação. Esta busca de ajuda poderá ocorrer durante a catalogação, ou então, antes da publicação dos *datasets*.

A seguir, estão apresentadas as orientações para o uso dos vocabulários controlados AGROVOC Thesaurus, NAL Thesaurus e NCBI Taxonomy, identificados junto à equipe do Projeto-piloto como apropriados à catalogação de *datasets* nos projetos-piloto para a implantação de Gestão de Dados de Pesquisa no Laboratório Multiusuário de Bioinformática, por serem considerados aderentes aos princípios FAIR, ou seja, tanto o AGROVOC, o NAL Thesaurus e o NCBI Taxonomy, quanto os seus termos são encontráveis, interoperáveis, acessíveis e reusáveis.

As orientações, a seguir, são destinadas especificamente para o preenchimento do elemento **Palavras-chave** nos campos **Termo-chave**, **Vocabulário controlado do termo-chave** e **URL do termo-chave**; bem como do elemento **Classificação** nos campos, **Termo de classificação**, **Vocabulário controlado da classificação** e **URL do termo de classificação**.

## AGROVOC Thesaurus

### O que é AGROVOC Thesaurus?

AGROVOC Thesaurus é um vocabulário controlado desenvolvido e mantido pela Food and Agriculture Organization of the United Nations (FAO), desde a década de 1980. Contém mais de 37 mil conceitos organizados de forma hierárquica, associativa e equivalente em 37 idiomas, incluindo o português europeu. Sua cobertura abrange extensa área de conhecimento como: alimentação, nutrição, agricultura, florestas, pesca, nomes científicos e comuns de animais e plantas, meio ambiente, aspectos biológicos e técnicas de cultivo de plantas e de criação de animais. (FAO, 2020).

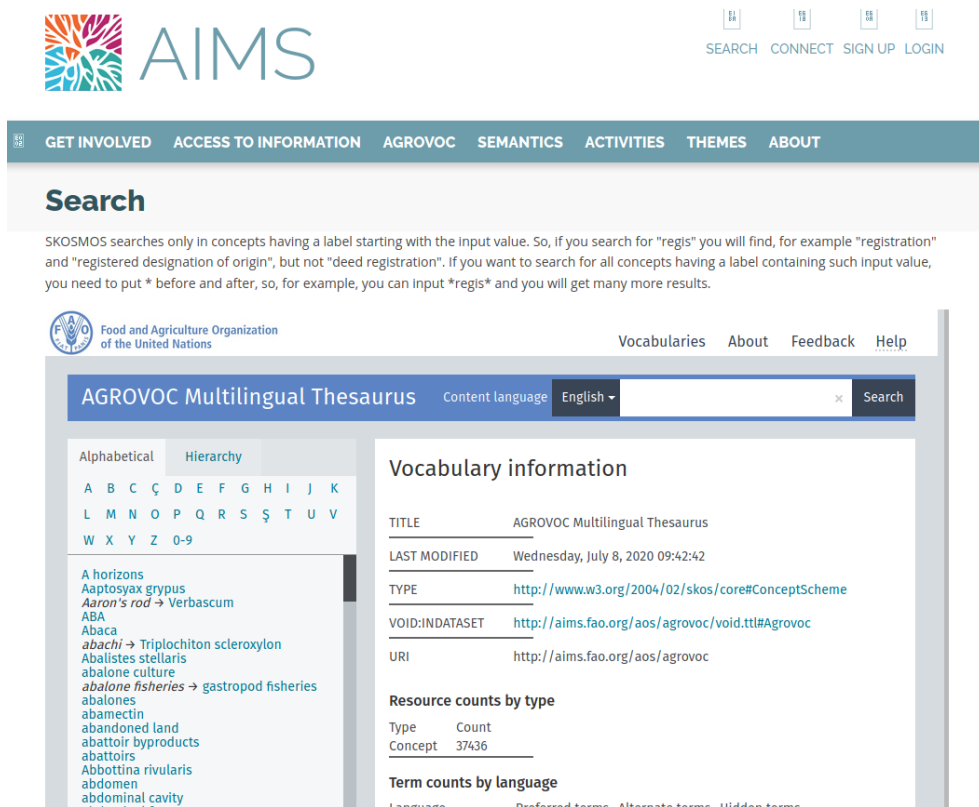
AGROVOC Thesaurus é amplamente usado por bibliotecas e repositórios digitais para indexar conteúdos, bem como para fins de organização de conhecimento. No âmbito do Repositório de Dados da Embrapa (AgDados), o AGROVOC Thesaurus é um dos vocabulários controlados escolhidos para descrever assunto (palavras-chave e classificação) de *dataset*.

### AGROVOC Thesaurus para Palavras-chave

A atribuição de palavras-chave é uma etapa da catalogação de *dataset* no AgDados, especificamente no preenchimento dos campos do Elemento **Palavras-chave**, no formulário de catalogação. Abaixo encontra-se o passo a passo para a realização de busca e o preenchimento dos campos **Termo-chave - Vocabulário controlado do termo-chave - URL do termo-chave**.

Seguem etapas do passo a passo:

**Etapas 1 - Acesso ao site do AGROVOC:** acesse o endereço: <<http://aims.fao.org/standards/agrovoc/functionalities/search>>. A página é apresentada como mostra a Figura 1.



The screenshot displays the AGROVOC Multilingual Thesaurus search interface. At the top, there is a navigation bar with links: GET INVOLVED, ACCESS TO INFORMATION, AGROVOC, SEMANTICS, ACTIVITIES, THEMES, and ABOUT. Below this is a search bar with the text "Search". A note explains that SKOSMOS searches only in concepts having a label starting with the input value, and provides examples of how to use wildcards like \* for more results.

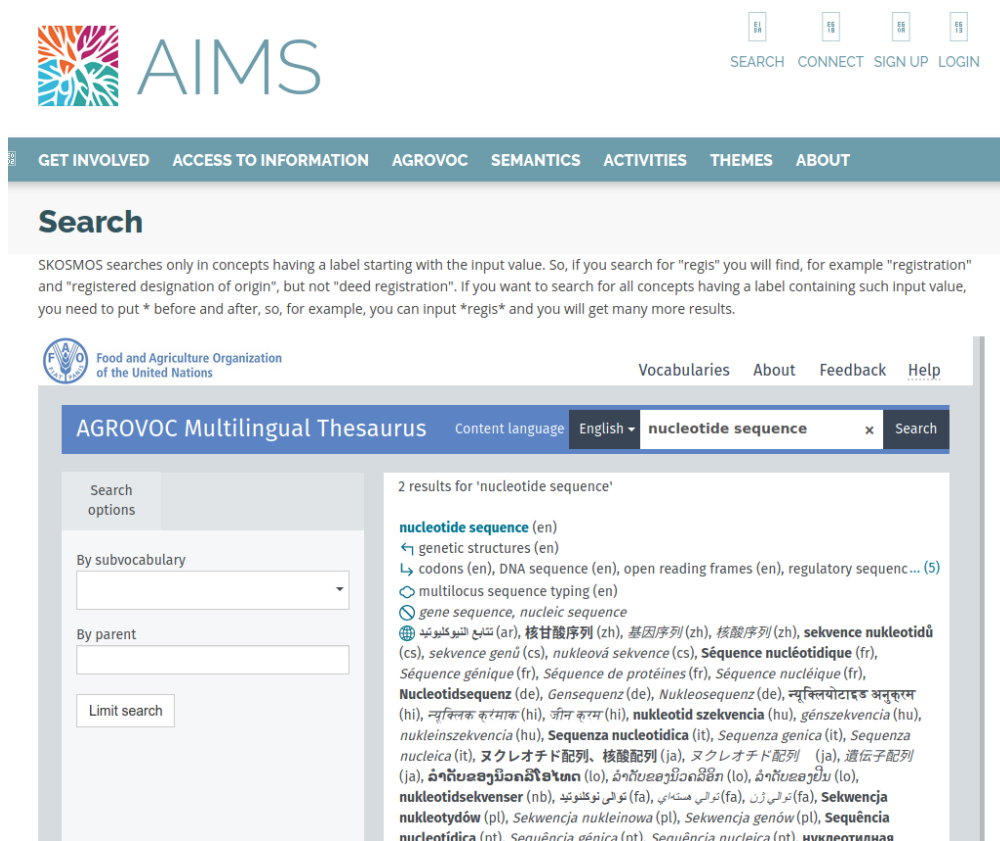
The main content area is titled "AGROVOC Multilingual Thesaurus" and includes a "Content language" dropdown set to "English" and a "Search" button. On the left, there is a sidebar with "Alphabetical" and "Hierarchy" tabs. The "Alphabetical" tab is active, showing a list of terms starting with 'A' through 'Z' and '0-9'. The "Hierarchy" tab is also visible, showing a list of terms starting with 'A' through 'Z' and '0-9'.

The main content area displays "Vocabulary information" for the selected term, including fields for TITLE, LAST MODIFIED, TYPE, VOID:IN DATASET, and URI. Below this, there is a section for "Resource counts by type" and "Term counts by language".

Type	Count
Concept	37436

Figura 1. Tela de busca do AGROVOC Thesaurus.

**Etapla 2 - Realização da busca:** no menu à esquerda clique em *Alphabetical*, e em seguida no campo *Search* digite o termo desejado, por exemplo *nucleotide sequence*, e execute a busca clicando em *Search*. Após a realização da busca, serão apresentados os resultados, como mostra a Figura 2.



The screenshot shows the AGROVOC Multilingual Thesaurus interface. At the top, there is a logo for AIMS (Agriculture Information Management System) and navigation links: GET INVOLVED, ACCESS TO INFORMATION, AGROVOC, SEMANTICS, ACTIVITIES, THEMES, ABOUT. Below this is a search bar with the text "Search" and a description: "SKOSMOS searches only in concepts having a label starting with the input value. So, if you search for 'regis' you will find, for example 'registration' and 'registered designation of origin', but not 'deed registration'. If you want to search for all concepts having a label containing such input value, you need to put \* before and after, so, for example, you can input \*regis\* and you will get many more results."

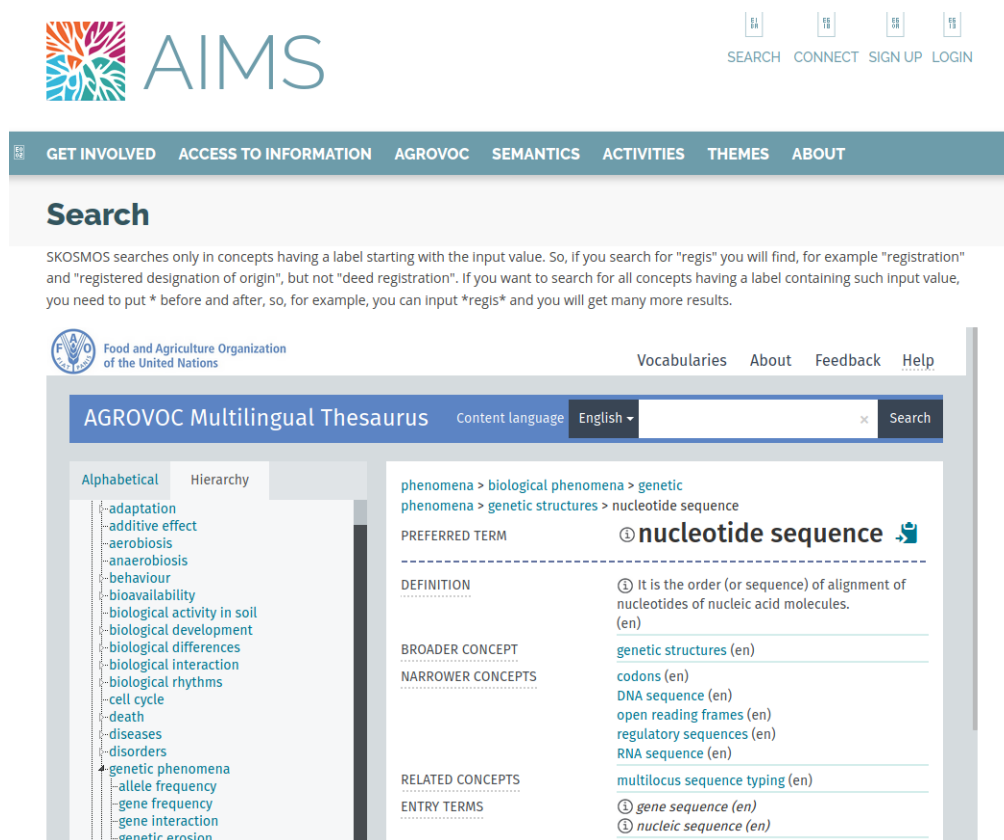
The main search area is titled "AGROVOC Multilingual Thesaurus" and includes a "Content language" dropdown set to "English" and a search input field containing "nucleotide sequence". Below the search bar, there are search options: "By subvocabulary" and "By parent", each with a dropdown menu, and a "Limit search" button.

The search results are displayed on the right side, showing "2 results for 'nucleotide sequence'". The results list includes:

- nucleotide sequence (en)**
- ↳ genetic structures (en)
- ↳ codons (en), DNA sequence (en), open reading frames (en), regulatory sequenc... (5)
- ↳ multilocus sequence typing (en)
- ↳ gene sequence, nucleic sequence
- ↳ تتابع النوكليوتيد (ar), 核苷酸序列 (zh), 基因序列 (zh), 核酸序列 (zh), **sekvence nukleotidů** (cs), **sekvence genů** (cs), **nukleová sekvence** (cs), **Séquence nucléotidique** (fr), **Séquence génique** (fr), **Séquence de protéines** (fr), **Séquence nucléique** (fr), **Nucleotidsequenz** (de), **Gensequenz** (de), **Nukleosequenz** (de), **न्यूक्लिओटाइड अनुक्रम** (hi), **न्यूक्लिक क्रमांक** (hi), **जीन क्रम** (hi), **nukleotid szekvencia** (hu), **génszekvencia** (hu), **nukleinszekvencia** (hu), **Sequenza nucleotidica** (it), **Sequenza genica** (it), **Sequenza nucleica** (it), **ヌクレオチド配列** (ja), **核酸配列** (ja), **ヌクレオチド配列** (ja), **遺伝子配列** (ja), **မာ်တီယဆေ့ၤမိၤဝဲၤမိၤမိၤမိၤ** (lo), **မာ်တီယဆေ့ၤမိၤဝဲၤမိၤမိၤမိၤ** (lo), **မာ်တီယဆေ့ၤမိၤဝဲၤမိၤမိၤမိၤ** (lo), **nukleotidsekvenser** (nb), **ترافى نوكليوتيد** (fa), **ترافى مستعاني** (fa), **ترافى ژن** (fa), **Sekwencja nukleotydów** (pl), **Sekwencja nukleinowa** (pl), **Sekwencja genów** (pl), **Sequência nucleotídica** (pt), **Sequência génica** (pt), **Sequência nucleica** (pt), **нуклеотидная**

**Figura 2.** Tela de resultado de busca do AGROVOC Thesaurus para o termo *nucleotide sequence*.


**Etapa 3 - Acesso a página detalhada do termo:** clique no termo *nucleotide sequence* no resultado da busca. A Figura 3 apresenta os detalhes do termo pesquisado.



The screenshot shows the AIMS website interface. At the top, there is a navigation bar with links: GET INVOLVED, ACCESS TO INFORMATION, AGROVOC, SEMANTICS, ACTIVITIES, THEMES, and ABOUT. Below this is a search bar with the text "Search".

The main content area is titled "AGROVOC Multilingual Thesaurus". It features a sidebar with a list of terms under the "Alphabetical" tab. The main content area displays the details for the term "nucleotide sequence".

**phenomena > biological phenomena > genetic phenomena > genetic structures > nucleotide sequence**

**PREFERRED TERM** ① **nucleotide sequence** 

**DEFINITION**  
① It is the order (or sequence) of alignment of nucleotides of nucleic acid molecules. (en)


**BROADER CONCEPT**  
genetic structures (en)

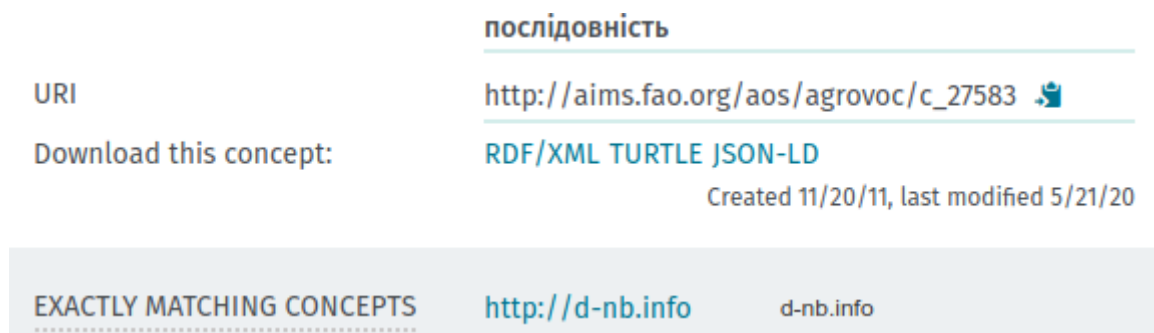
**NARROWER CONCEPTS**  
codons (en)  
DNA sequence (en)  
open reading frames (en)  
regulatory sequences (en)  
RNA sequence (en)

**RELATED CONCEPTS**  
multilocus sequence typing (en)

**ENTRY TERMS**  
① gene sequence (en)  
① nucleic sequence (en)

**Figura 3.** Tela do termo do *nucleotide sequence* contendo informações detalhadas.

**Etapa 4 - Obtenção da URL do termo:** na página com os detalhes do termo, apresentado na Figura 3, deslize o mouse para o fim da tela, e procure o item URI, como mostrado na Figura 4. Copie o link apresentado, ou então, clique no ícone  para copiar e levar o endereço para o formulário de catalogação.



**Figura 4.** Resultado da busca do termo *nucleotide sequence*, no AGROVOC Thesaurus, com destaque para o item URI.

**Etapa 5: Preenchimento dos campos do elemento Palavras-chave:** no formulário de catalogação do *dataset* no AgDados, preencha os campos conforme o exemplo:

**Palavra-chave:** nucleotide sequence

**Vocabulário controlado do termo-chave:** AGROVOC Thesaurus

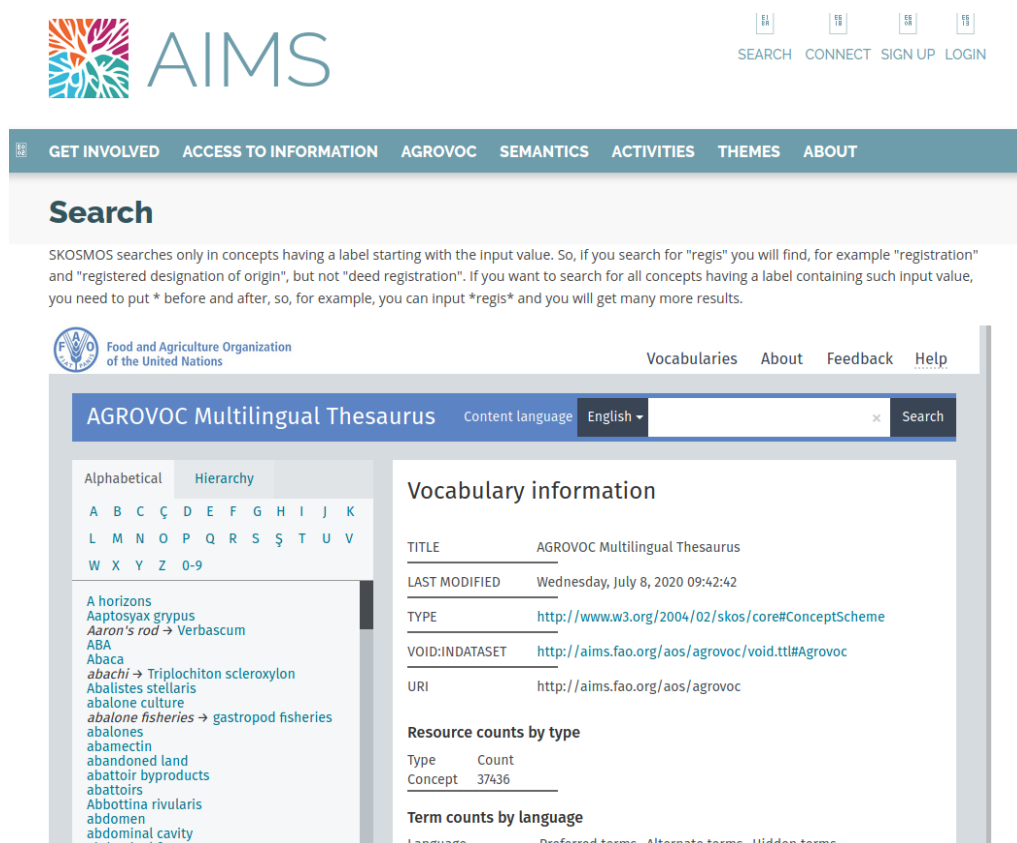
**URL da palavra-chave:** [http://aims.fao.org/aos/agrovoc/c\\_27583](http://aims.fao.org/aos/agrovoc/c_27583)

## AGROVOC Thesaurus para Classificação

A classificação do assunto do *dataset* é feita por meio da atribuição de uma ou mais categorias de assunto. Tais categorias representam estruturas de assunto mais abrangentes sob as quais um assunto específico, como por exemplo, *nucleotide sequence*, está inserido. A atribuição de classificação de assunto, por meio de categorias, é uma etapa da catalogação do *dataset*, especificamente no preenchimento do campo **Termo de classificação**, do Elemento Classificação, no formulário de catalogação. Abaixo encontra-se o passo a passo para a realização de busca e o preenchimento dos campos **Termo de classificação** - **Vocabulário controlado da classificação** - **URL do termo de classificação**.

Seguem etapas do passo a passo:

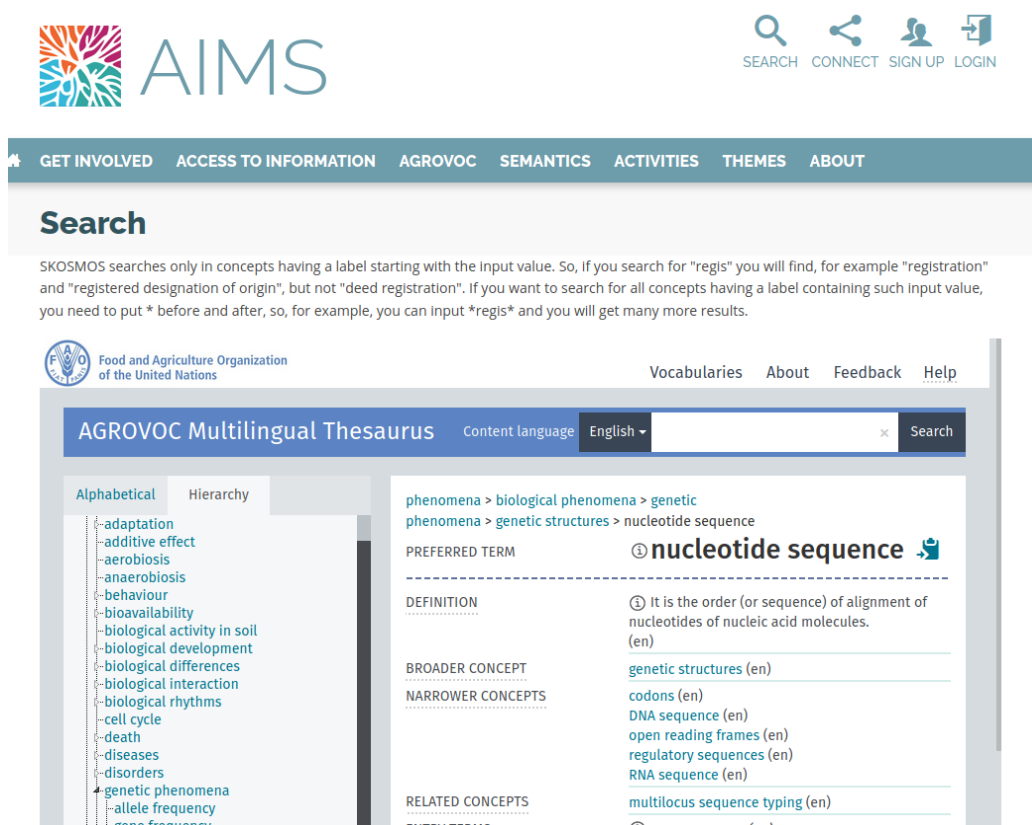
**Etapa 1 - Acesso ao site do AGROVOC:** acesse o endereço: <http://aims.fao.org/standards/agrovoc/functionalities/search>. A página é apresentada como mostra a Figura 5.



The screenshot shows the AGROVOC Multilingual Thesaurus search interface. At the top, there's a navigation bar with links: GET INVOLVED, ACCESS TO INFORMATION, AGROVOC, SEMANTICS, ACTIVITIES, THEMES, ABOUT. Below this is a search bar with the text "Search". To the right of the search bar are links: SEARCH, CONNECT, SIGN UP, LOGIN. Below the search bar, there's a section titled "Search" with a description of SKOSMOS searches. The main content area displays "AGROVOC Multilingual Thesaurus" with a language dropdown set to "English". It includes a "Vocabulary information" section with details like title, last modified date, type, void:indataset, and uri. There are also sections for "Resource counts by type" and "Term counts by language".

Figura 5. Tela de busca do AGROVOC Thesaurus.

**Etapa 2 - Realização da busca:** no menu à esquerda clique em *Hierarchy*, em seguida no campo *Search* digite o termo desejado para o qual necessita fazer a classificação de assunto do *dataset*, por exemplo, *nucleotide sequence*, e execute a busca clicando em *Search*. Após a realização da busca, serão apresentados os resultados, como mostra a Figura 6.

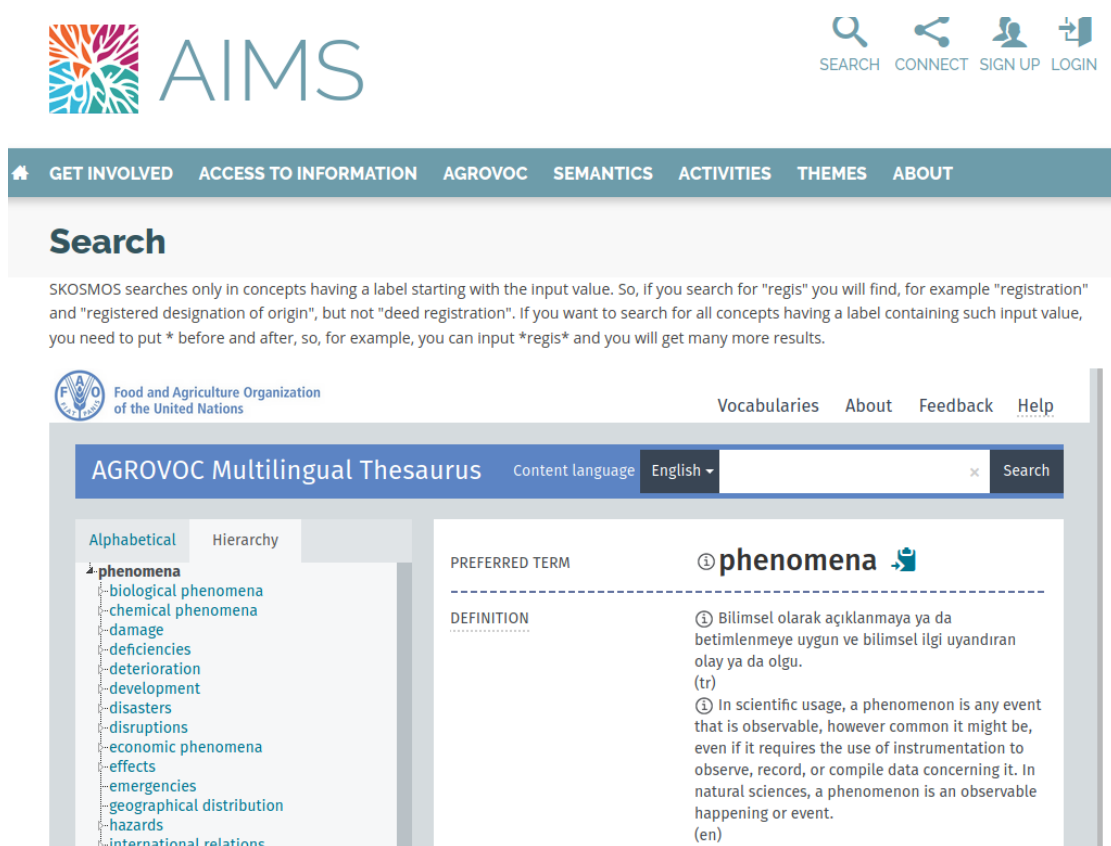


The screenshot shows the AGROVOC Multilingual Thesaurus interface. At the top, there is a navigation bar with links: GET INVOLVED, ACCESS TO INFORMATION, AGROVOC, SEMANTICS, ACTIVITIES, THEMES, and ABOUT. Below this is a search bar with the text "Search". A note explains that SKOSMOS searches only in concepts having a label starting with the input value. Below the note is the FAO logo and the text "Food and Agriculture Organization of the United Nations". The main header of the thesaurus is "AGROVOC Multilingual Thesaurus" with a language dropdown set to "English" and a "Search" button. The left sidebar shows a list of concepts under the "Hierarchy" tab, with "genetic phenomena" selected. The main content area displays the search results for "nucleotide sequence". It shows the preferred term "nucleotide sequence" with a definition: "It is the order (or sequence) of alignment of nucleotides of nucleic acid molecules. (en)". It also lists broader concepts (genetic structures (en)), narrower concepts (codons (en), DNA sequence (en), open reading frames (en), regulatory sequences (en), RNA sequence (en)), and related concepts (multilocus sequence typing (en)).

**Figura 6.** Tela de resultado de busca do AGROVOC Thesaurus para o termo *nucleotide sequence*.



**Etapa 3 - Busca das categorias:** na tela de resultados apresentada na Figura 6, observe logo acima do termo *nucleotide sequence*, uma sequência de termos que representam as suas categorias hierárquicas, sendo elas: *phenomena* > *biological phenomena* > *genetic phenomena* > *genetic structures* > *nucleotide sequence*. Essas categorias são as classificações de assunto encontradas para o termo *nucleotide sequence*. Caberá ao especialista do domínio orientar a escolha da(s) categoria(s) que melhor classifica(m) o assunto do *dataset*. Ao escolher a categoria *phenomena*, por exemplo, clique este mesmo termo para abrir sua página, como mostra a Figura 7.



The screenshot displays the AIMS AGROVOC Multilingual Thesaurus interface. At the top, there is a navigation bar with links: GET INVOLVED, ACCESS TO INFORMATION, AGROVOC, SEMANTICS, ACTIVITIES, THEMES, and ABOUT. Below this is a search bar with the text "Search" and a description of the search functionality. The main content area is divided into two sections: "Alphabetical" and "Hierarchy". The "Alphabetical" section lists various terms, including "phenomena", "biological phenomena", "chemical phenomena", "damage", "deficiencies", "deterioration", "development", "disasters", "disruptions", "economic phenomena", "effects", "emergencies", "geographical distribution", "hazards", and "international relations". The "Hierarchy" section shows the preferred term "phenomena" with its definition in multiple languages. The definition is provided in Turkish (tr) and English (en).

**Search**

SKOSMOS searches only in concepts having a label starting with the input value. So, if you search for "regis" you will find, for example "registration" and "registered designation of origin", but not "deed registration". If you want to search for all concepts having a label containing such input value, you need to put \* before and after, so, for example, you can input \*regis\* and you will get many more results.

Food and Agriculture Organization of the United Nations

Vocabularies About Feedback Help

AGROVOC Multilingual Thesaurus Content language English Search

Alphabetical Hierarchy

phenomena

- biological phenomena
- chemical phenomena
- damage
- deficiencies
- deterioration
- development
- disasters
- disruptions
- economic phenomena
- effects
- emergencies
- geographical distribution
- hazards
- international relations

PREFERRED TERM


phenomena


DEFINITION

① Bilimsel olarak açıklanmaya ya da betimlenmeye uygun ve bilimsel ilgi uyandıran olay ya da olgu.  
(tr)

① In scientific usage, a phenomenon is any event that is observable, however common it might be, even if it requires the use of instrumentation to observe, record, or compile data concerning it. In natural sciences, a phenomenon is an observable happening or event.  
(en)

Figura 7. Tela com os detalhes do termo *phenomena*.

**Etapa 4: Obtenção da URL para o termo:** na página com os detalhes do termo, apresentado na Figura 7, deslize o mouse para o fim da tela, e procure o item URI, como mostrado na Figura 8. Copie o link apresentado, ou então clique no ícone  para copiar e levar o endereço para o formulário de catalogação.

URI	<a href="http://aims.fao.org/aos/agrovoc/c_330704">http://aims.fao.org/aos/agrovoc/c_330704</a> 
Download this concept:	<a href="#">RDF/XML</a> <a href="#">TURTLE</a> <a href="#">JSON-LD</a>

**Figura 8.** Resultado da busca do termo *phenomena*, no AGROVOC Thesaurus, com destaque para o item URI.

**Etapa 5: Preenchimento dos campos do elemento Classificação:** no formulário de catalogação do *dataset* no AgDados, preencha os campos conforme o exemplo:

Termo de classificação: *phenomena*

Vocabulário controlado da classificação: AGROVOC Thesaurus

URL do termo de classificação: [http://aims.fao.org/aos/agrovoc/c\\_330704](http://aims.fao.org/aos/agrovoc/c_330704)

## NAL Thesaurus

### O que é NAL Thesaurus?

NAL Thesaurus é um vocabulário controlado desenvolvido e mantido pela National Agricultural Library dos Estados Unidos. Lançado em 2002, este vocabulário bilíngue (inglês e espanhol) tem abrangência ampla em agricultura, biologia e disciplinas correlatas. Contém mais 260 mil termos, que são organizados em 17 categorias de assunto. Além disso, inclui relações hierárquicas, associativas e de equivalência entre conceitos. (Estados Unidos, 2020).

NAL Thesaurus é usado principalmente para indexação de assunto de diferentes objetos de informação, dentre os quais, os *datasets*. No âmbito do Repositório de Dados da Embrapa (AgDados), o NAL Thesaurus é um dos vocabulários controlados escolhidos para descrever as palavras-chave e classificação do *dataset*.

### NAL Thesaurus para Palavras-chave

A atribuição de palavras-chave é uma etapa da catalogação de dataset no AgDados, especificamente no preenchimento dos campos do Elemento **Palavras-chave**, no formulário de catalogação. Abaixo encontra-se o passo a passo para a realização de busca e o preenchimento dos campos **Termo-chave - Vocabulário controlado do termo-chave - URL do termo-chave**.

Seguem etapas do passo a passo:

**Etapa 1 - Acesso ao site do NAL Thesaurus:** acesse o endereço:

<https://agclass.nal.usda.gov/dne/search.shtml>. A página é apresentada como mostra a Figura 9.

USDA United States Department of Agriculture  
National Agricultural Library

Inter-American Institute for Cooperation on Agriculture IICA

tesauro thesaurus glosario glossary

Home About Thesaurus and Glossary Download Files Help Contact Us En Español

You are here: [Home](#) / [Search Thesaurus](#)

## Thesaurus: Search

Search term or text to match:

Select English or Spanish:

Select a search method:

Number of terms to display:

[Back to Top](#)

Last Modified: Dec 15, 2017

I Want To...

- [Search in Spanish](#)
- [Download Thesaurus File](#)
- [Search the NAL Catalog \(AGRICOLA\)](#)

NAL | IICA | USDA | Agricultural Research Service | Science.gov | Web Policies and Important Links | Site Map  
FOIA | Accessibility Statement | Privacy Policy | Non-Discrimination Statement | Information Quality | USA.gov | WhiteHouse.gov

**Figura 9.** Tela de busca do NAL Thesaurus.

**Etapa 2 - Realização da busca:** no campo *Search term* digite o termo desejado. O termo *mites* foi escolhido como exemplo. No campo *Select a search method* escolha a opção mais adequada. No exemplo mostrado foi escolhida a opção *Terms start with*. Clique no botão *Search*. A tela de busca é apresentada na Figura 10 e os resultados na Figura 11.

USDA United States Department of Agriculture National Agricultural Library

Inter-American Institute for Cooperation on Agriculture IICA

tesauro thesaurus glosario glossary

Home About Thesaurus and Glossary Download Files Help Contact Us En Español

You are here: Home / Search Thesaurus

### Thesaurus: Search

Search term or text to match:

Select English or Spanish:  
English | inglés

Select a search method:  
Terms start with ...

Number of terms to display:  
200

Search Reset

Back to Top

Last Modified: Dec 15, 2017

I Want To...

- Search in Spanish
- Download Thesaurus File
- Search the NAL Catalog (AGRICOLA)

NAL | IICA | USDA | Agricultural Research Service | Science.gov | Web Policies and Important Links | Site Map  
FOIA | Accessibility Statement | Privacy Policy | Non-Discrimination Statement | Information Quality | USA.gov | WhiteHouse.gov

**Figura 10.** Tela de busca do NAL Thesaurus com o termo *mites*.



Figura 11. Resultado da busca do termo no NAL Thesaurus.

**Etapa 3 - Acesso a página detalhada do termo:** clique no termo *mites* no resultado da busca. Um conjunto de informações adicionais será apresentado, conforme a Figura 12.

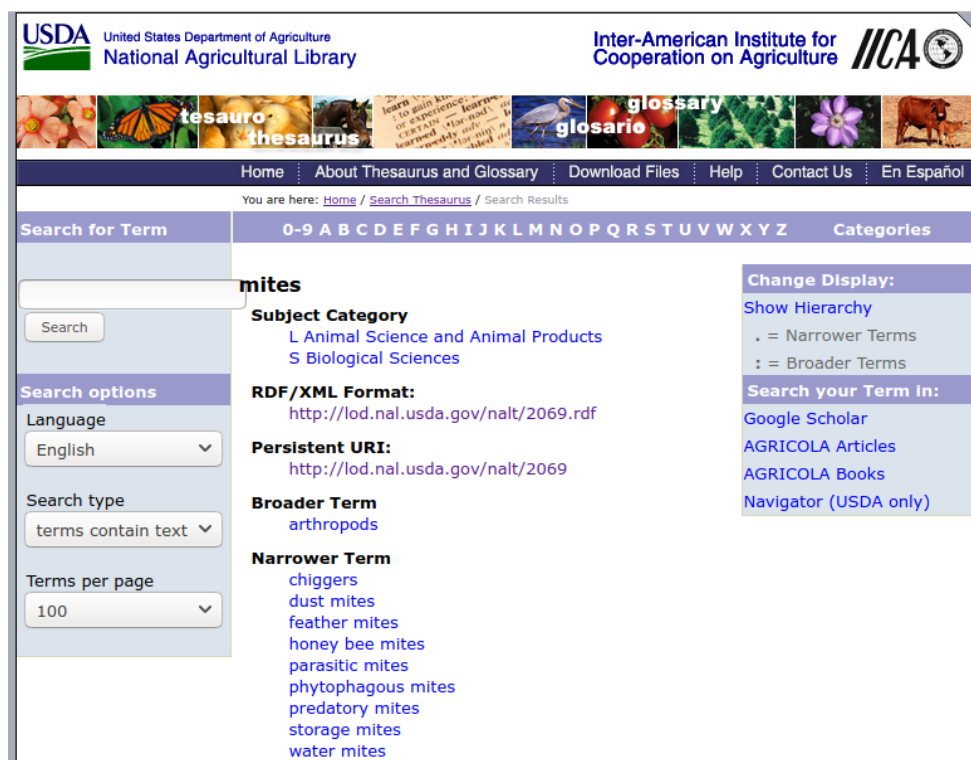


Figura 12. Tela do termo 'mites' contendo informações sobre classificações, relações, associações, formato, identificador persistente, tradução do termo para o espanhol e número do termo.

**Etapa 4 - Obtenção da URL do termo:** na página do termo *mites*, apresentada na Figura 12 copie a URL diretamente do endereço da página no navegador, e cole-a no campo **URL do termo-chave** do Elemento Palavras-chave, do formulário de catalogação do Repositório.

**Etapa 5: Preenchimento dos campos do elemento Palavras-chave:** no formulário de catalogação do *dataset* no Repositório de Dados, preencha os campos conforme o exemplo:

Termo-chave: mites

Vocabulário controlado do termo-chave: NAL Thesaurus

URL do termo-chave:

<https://agclass.nal.usda.gov/mtwdk.exe?k=default&l=60&w=2069&s=5&t=2>

## NAL Thesaurus para Classificação

A classificação do assunto do *dataset* é feita por meio da atribuição de uma ou mais categorias de assunto. Tais categorias representam estruturas de assunto mais abrangentes sob as quais um assunto específico, como por exemplo, 'mites', está inserido. A atribuição de classificação de assunto, por meio de categorias, é uma etapa da catalogação do *dataset*, especificamente no preenchimento do campo **Termo de classificação**, do Elemento Classificação no formulário de catalogação. Abaixo encontra-se o passo a passo para a realização de busca e o preenchimento dos campos **Termo de classificação - Vocabulário controlado da classificação - URL do termo de classificação**.

Seguem etapas do passo a passo:

**Etapas 1 - Acesso ao site do NAL Thesaurus:** acesse o endereço: <<https://agclass.nal.usda.gov/dne/search.shtml>>. A página é apresentada como mostra a Figura 13.

The screenshot displays the NAL Thesaurus search interface. At the top, the USDA logo and 'United States Department of Agriculture National Agricultural Library' are on the left, while the IICA logo and 'Inter-American Institute for Cooperation on Agriculture' are on the right. Below this is a navigation bar with links: Home, About Thesaurus and Glossary, Download Files, Help, Contact Us, and En Español. A banner with various agricultural images and the words 'tesauro' and 'glossario' is positioned above the navigation bar. The main content area has a green header that says 'Thesaurus: Search'. On the left, a sidebar titled 'Thesaurus' contains links for 'Search Thesaurus', 'Browse A-Z', and 'Browse by Subject Category'. Below this, a 'Glossary' section has links for 'Search Glossary' and 'Browse A-Z'. The main search area includes a text input field for 'Search term or text to match:', a dropdown for 'Select English or Spanish:' (currently set to 'English | inglés'), a dropdown for 'Select a search method:' (currently set to 'Terms which contain this character string'), and a dropdown for 'Number of terms to display:' (currently set to '200'). There are 'Search' and 'Reset' buttons at the bottom of the search area. On the right, a box titled 'I Want To...' contains links for 'Search in Spanish', 'Download Thesaurus File', and 'Search the NAL Catalog (AGRICOLA)'. At the bottom of the page, there is a 'Back to Top' link and a 'Last Modified: Dec 15, 2017' timestamp. The footer contains a list of links: NAL | IICA | USDA | Agricultural Research Service | Science.gov | Web Policies and Important Links | Site Map | FOIA | Accessibility Statement | Privacy Policy | Non-Discrimination Statement | Information Quality | USA.gov | WhiteHouse.gov.

**Figura 13.** Tela de busca do NAL Thesaurus.



**Etapa 2 - Realização da busca:** no campo *Search term* digite o termo desejado para o qual pretende realizar a classificação de assunto do *dataset*. Usaremos o termo *mites* como exemplo. Escolha a opção *Select a search method* mais adequada. No exemplo usaremos a opção *Terms start with*. Clique no botão *Search*. A tela de busca é apresentada na Figura 14 e os resultados na Figura 15.

The screenshot displays the NAL Thesaurus search page. At the top, there are logos for USDA (United States Department of Agriculture, National Agricultural Library) and IICA (Inter-American Institute for Cooperation on Agriculture). Below the logos is a navigation bar with links: Home, About Thesaurus and Glossary, Download Files, Help, Contact Us, and En Español. The main content area is titled 'Thesaurus: Search'. On the left, there is a sidebar with 'Thesaurus' and 'Glossary' sections. The 'Thesaurus' section has links for 'Search Thesaurus', 'Browse A-Z', and 'Browse by Subject Category'. The 'Glossary' section has links for 'Search Glossary' and 'Browse A-Z'. The main search area contains a 'Search term or text to match:' field with the value 'mites'. Below this is a 'Select English or Spanish:' dropdown menu set to 'English | inglés'. There is also a 'Select a search method:' dropdown menu set to 'Terms start with ...'. A 'Number of terms to display:' dropdown menu is set to '200'. At the bottom of the search area are 'Search' and 'Reset' buttons. On the right side, there is a 'I Want To...' section with links: 'Search in Spanish', 'Download Thesaurus File', and 'Search the NAL Catalog (AGRICOLA)'. At the bottom of the page, there is a 'Back to Top' link and a 'Last Modified: Dec 15, 2017' timestamp. The footer contains a list of links: NAL | IICA | USDA | Agricultural Research Service | Science.gov | Web Policies and Important Links | Site Map | FOIA | Accessibility Statement | Privacy Policy | Non-Discrimination Statement | Information Quality | USA.gov | WhiteHouse.gov.

**Figura 14.** Tela de busca do NAL Thesaurus para o termo *mites*.

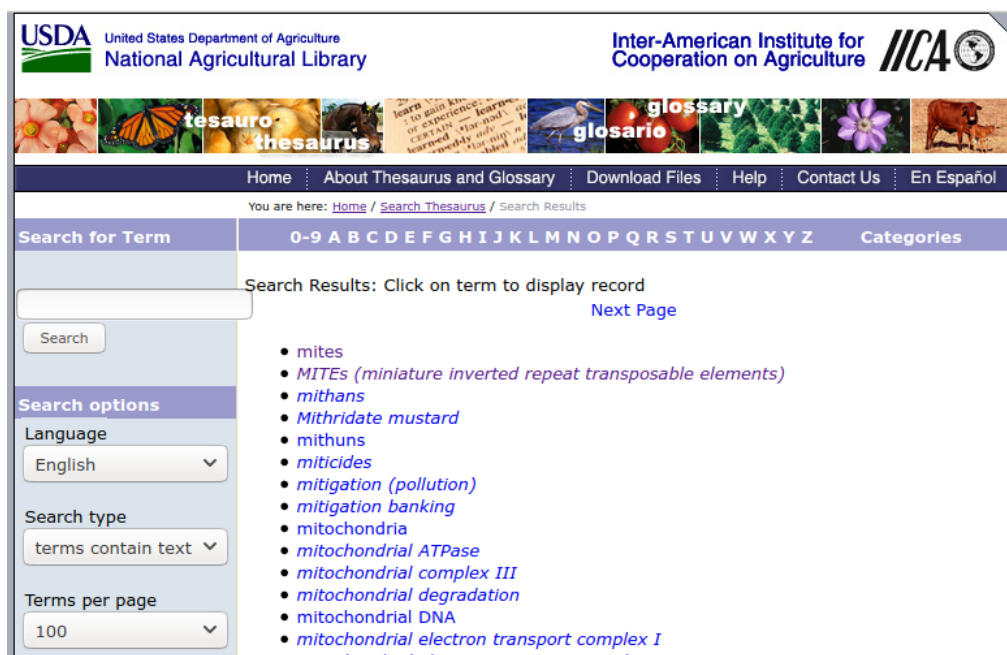


Figura 15. Resultado da busca do termo *mites* no NAL Thesaurus.

**Etapas 3 - Acesso a página detalhada do termo:** clique no termo *mites* no resultado da busca.

Um conjunto de informações adicionais será apresentado, conforme a Figura 16.

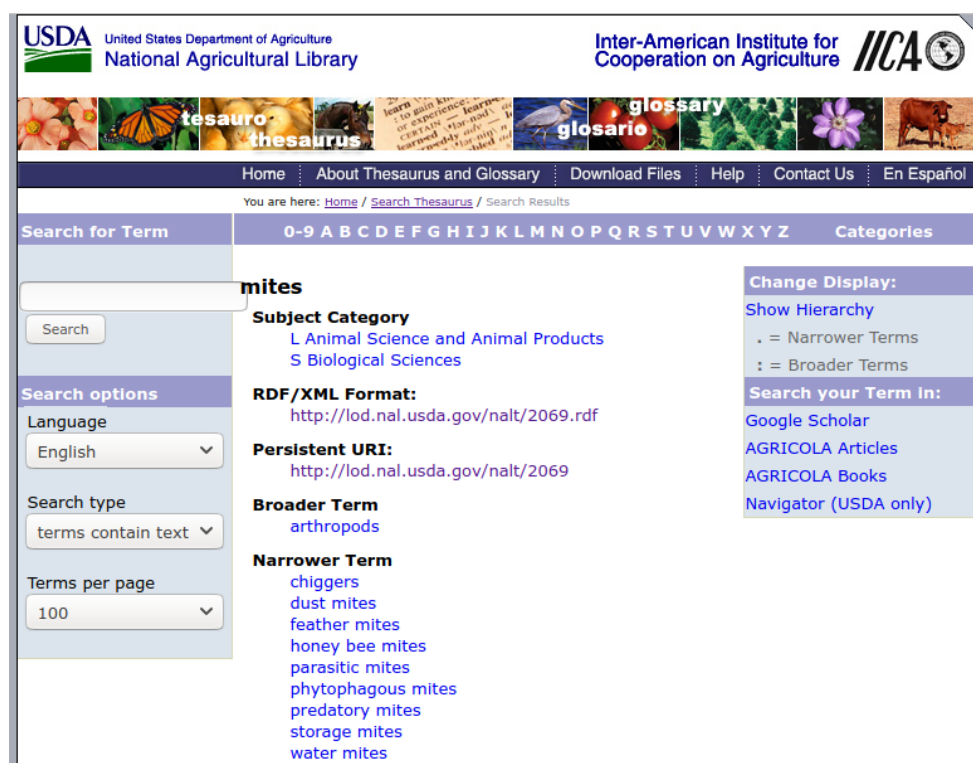
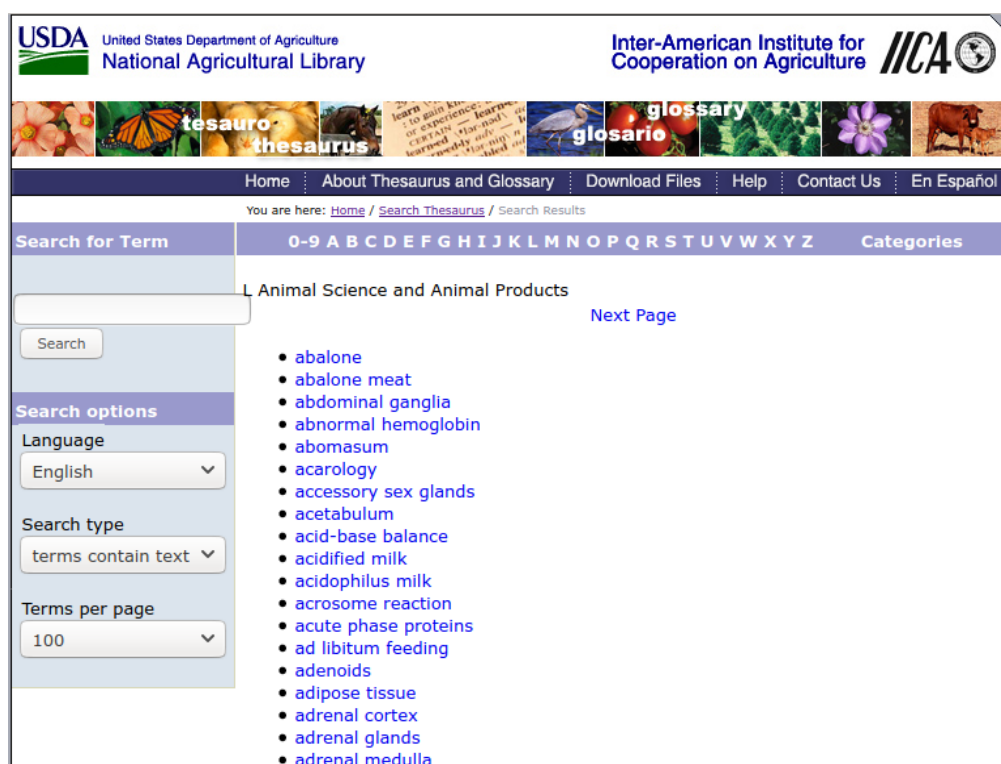


Figura 16. Tela do termo *mites* contendo informações sobre classificações, relações, associações, formato, identificador persistente, tradução do termo para o espanhol e número do termo.

**Etapa 4 - Obtenção da URL para o termo:** na página do termo *mites*, apresentado na Figura 16, escolha em *Subject Category* uma categoria de assunto que melhor represente o termo *mites*, como por exemplo *L Animal Science and Animal Product*. Clique no item selecionado. A página da figura 17 será apresentada. Copie a URL da página, que será a URL do termo de classificação.



**Figura 17.** Tela da categoria 'L Animal Sciences and Animal Products' do NAL Thesaurus.

**Etapa 5: Preenchimento dos campos do elemento Classificação:** no formulário de catalogação do *dataset* no AgDados, preencha os campos conforme o exemplo:

Termo de classificação: L Animal Science and Animal Products

Vocabulário controlado da classificação: NAL Thesaurus

URL do termo de classificação:

<https://agclass.nal.usda.gov/mtwdk.exe?k=default&l=60&s=cid&w=100>

## NCBI Taxonomy

### O que é NCBI Taxonomy?

National Center for Biotechnology Information (NCBI) Taxonomy é um sistema de informação que inclui nomes e classificações de organismos de sequências de nucleotídeos e de proteínas. NCBI Taxonomy é também o repositório padrão de nomenclatura e classificação para o International Nucleotide Sequence Database Collaboration (INSDC). Contém uma lista de nomes considerados nomenclaturalmente corretos ou válidos, os quais são definidos de acordo com diferentes códigos de nomenclatura. Inclui ainda vários nomes que existem a despeito de códigos. (National Center for Biotechnology Information, 2020). NCBI Taxonomy integra o Entrez, um metabuscador mantido pelo NCBI, que coleta, trata e disponibiliza informação relevante para o desenvolvimento da biotecnologia.

NCBI Taxonomy é usado principalmente para indexação de assunto de diferentes objetos de informação, dentre os quais, os *datasets*. No âmbito do Repositório de Dados da Embrapa (AgDados), o NCBI é um dos vocabulários controlados escolhidos para descrever assunto (palavras-chave e classificação) de *datasets*.

### NCBI Taxonomy para Palavras-chave

A atribuição de palavras-chave é uma etapa da catalogação de *dataset* no AgDados, especificamente no preenchimento dos campos do Elemento **Palavras-chave**, no formulário de catalogação. Abaixo encontra-se o passo a passo para a realização de busca e o preenchimento dos campos **Termo-chave - Vocabulário controlado do termo-chave - URL do termo-chave**.

Seguem etapas do passo a passo:

**Etapa 1 - Acesso ao site do NCBI Taxonomy:** acesse o endereço: <<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>>. A página é apresentada como mostra a Figura 18.



The "Token set" option returns longer names that include the search terms "Complete match" option versus the "Set of tokens" option. The "Phonetic : organism name. It tries to find the phonetically closest strings (try "Drozofi

**This is the top level of the taxonomy database maintained by NCBI/GenBank. Yo**

- [Archaea](#)
- [Bacteria](#)
- [Eukaryota](#)
- [Viruses](#)
- [Other](#)
- [Unclassified](#)

---

**Figura 18.** Tela de busca do NCBI Taxonomy.

**Etapa 2 - Realização da busca:** no campo *Search for* digite o termo desejado, como por exemplo *Rhipicephalus (Boophilus) microplus*. Clique em *Go*, e veja os resultados obtidos. A Figura 19 apresenta os resultados.

The screenshot shows the NCBI Taxonomy Browser interface. At the top, there are logos for NCBI and various biological entities (mushroom, virus, fish, flower). Below the logos, there are tabs for Entrez, PubMed, Nucleotide, Protein, and Genome. The search bar contains the text "Rhipicephalus (Boophilus) microplus" and the dropdown menu is set to "complete name". There is a "lock" checkbox checked. Below the search bar, there are fields for "Display" (set to 3) and "levels using filter" (set to none). The search results are displayed below the search bar, showing the taxonomic information for *Rhipicephalus microplus*.

**Rhipicephalus microplus**

Taxonomy ID: 6941 (for references in articles please use NCBI:txid6941)

current name

***Rhipicephalus microplus*** (Canestrini, 1888)

homotypic genbank synonym: ***Boophilus microplus***

homotypic synonym: ***Rhipicephalus (Boophilus) microplus***

Genbank common name: **southern cattle tick**

NCBI BLAST name: **mites & ticks**

Rank: **species**

Genetic code: [Translation table 1 \(Standard\)](#)

Mitochondrial genetic code: [Translation table 5 \(Invertebrate Mitochondrial\)](#)

Other names:

common name(s)

**cattle tick**

**Figura 19.** Tela de resultados da busca do termo *Rhipicephalus (Boophilus) microplus* no NCBI Taxonomy.

**Etapa 3 - Obtenção da URL para o termo:** na página de resultado de busca do termo *Rhipicephalus microplus*, apresentado na Figura 19, no frame à direita, em *Taxonomy*, clique no endereço correspondente na coluna *Direct links* para abrir a página do referido termo. Em seguida, clique em *Rhipicephalus microplus* para abrir a página deste termo. Copie a URL desta página e cole-a no campo **URL do termo-chave** do Elemento Palavras-chave, do formulário de catalogação do Repositório.

**Etapa 4: Preenchimento dos campos do elemento Palavras-chave:** no formulário de catalogação do *dataset* no AgDados, preencha os campos conforme o exemplo:

Termo-chave: *Rhipicephalus microplus*

Vocabulário controlado do termo-chave: NCBI Taxonomy


URL do termo-chave: <https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=6941>

## NCBI Taxonomy para Classificação

A classificação do assunto do *dataset* é feita por meio da atribuição de uma ou mais categorias de assunto. Tais categorias representam estruturas de assunto mais abrangentes sob as quais um assunto específico, como por exemplo, '*Rhipicephalus (Boophilus) microplus*', está inserido. A atribuição de classificação de assunto, por meio de categorias, é uma etapa da catalogação do *dataset*, especificamente no preenchimento do campo **Termo de classificação**, do Elemento Classificação no formulário de catalogação. Abaixo encontra-se o passo a passo para a realização de busca e o preenchimento dos campos **Termo de classificação - Vocabulário - URL do termo de classificação**.

Seguem etapas do passo a passo:

**Etapa 1 - Acesso ao site do NCBI Taxonomy:** acesse o endereço: <<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>>. A página é apresentada como mostra a Figura 20.



The "Token set" option returns longer names that include the search terms "Complete match" option versus the "Set of tokens" option. The "Phonetic" option returns the phonetic organism name. It tries to find the phonetically closest strings (try "Drozofi

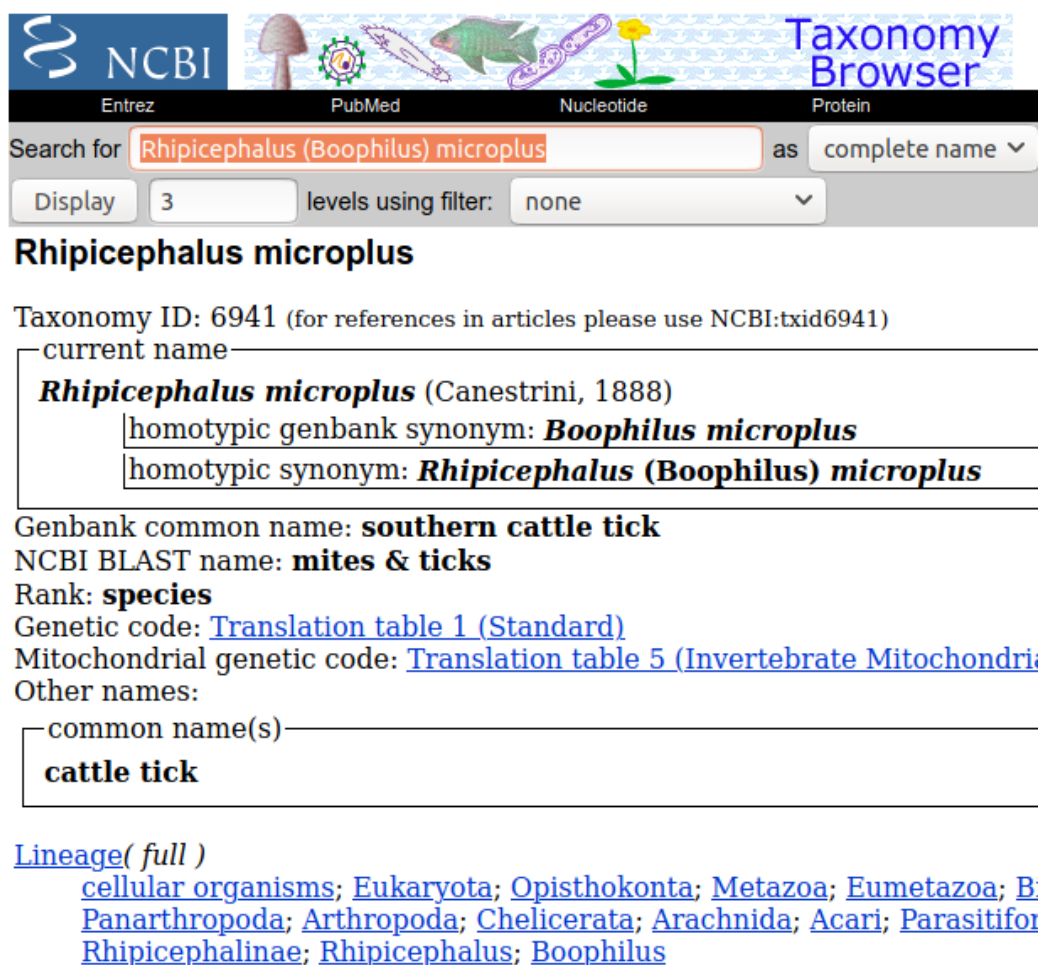
**This is the top level of the taxonomy database maintained by NCBI/GenBank. You**

- [Archaea](#)
- [Bacteria](#)
- [Eukaryota](#)
- [Viruses](#)
- [Other](#)
- [Unclassified](#)

**Figura 20.** Tela de busca do NCBI Taxonomy.



**Etapa 2 - Realização da busca:** no campo *Search for* digite o termo desejado para o qual pretende fazer a classificação de assunto, como por exemplo *Rhipicephalus (Boophilus) microplus*. Clique em *Go*, e veja os resultados obtidos. A Figura 21 apresenta os resultados.



The image shows the NCBI Taxonomy Browser interface. At the top, there are logos for NCBI and various organisms (mushroom, virus, fish, bacteria, flower). Below the logos are tabs for Entrez, PubMed, Nucleotide, and Protein. The search bar contains the text "Rhipicephalus (Boophilus) microplus" and the dropdown menu is set to "complete name". The "Display" button is set to "3" and the "levels using filter" is set to "none".

**Rhipicephalus microplus**

Taxonomy ID: 6941 (for references in articles please use NCBI:txid6941)

current name

***Rhipicephalus microplus*** (Canestrini, 1888)

homotypic genbank synonym: ***Boophilus microplus***

homotypic synonym: ***Rhipicephalus (Boophilus) microplus***

Genbank common name: **southern cattle tick**

NCBI BLAST name: **mites & ticks**

Rank: **species**

Genetic code: [Translation table 1 \(Standard\)](#)

Mitochondrial genetic code: [Translation table 5 \(Invertebrate Mitochondrial\)](#)

Other names:

common name(s)

**cattle tick**

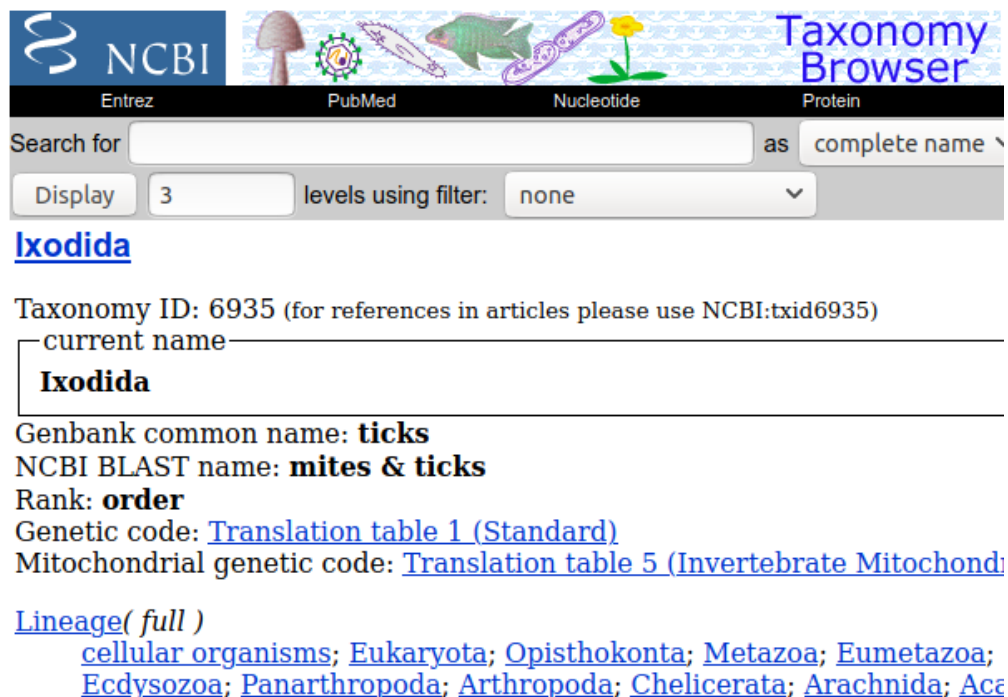
[Lineage\( full \)](#)

[cellular organisms](#); [Eukaryota](#); [Opisthokonta](#); [Metazoa](#); [Eumetazoa](#); [B](#); [Panarthropoda](#); [Arthropoda](#); [Chelicerata](#); [Arachnida](#); [Acari](#); [Parasitiformes](#); [Rhipicephalinae](#); [Rhipicephalus](#); [Boophilus](#)

**Figura 21.** Tela de resultados da busca do termo *Rhipicephalus (Boophilus) microplus* no NCBI Taxonomy.



**Etapa 3 - Seleção do termo classificação:** no resultado de busca mostrado na Figura 22, escolha em *Lineage (full)* uma categoria de assunto que melhor classifique o termo *Rhipicephalus microplus*, como por exemplo, *Ixodida*. A Figura 22 apresenta o conteúdo para o termo escolhido. Selecione o item *Ixodida*, e a Figura 23 será apresentada.



The screenshot shows the NCBI Taxonomy Browser interface. At the top, there are logos for NCBI, Entrez, PubMed, Nucleotide, and Protein. Below these is a search bar with the text "Search for" and a dropdown menu set to "as complete name". Below the search bar is a "Display" button and a "3 levels using filter: none" dropdown. The main content area shows the classification hierarchy for ***Ixodida***. The "current name" is ***Ixodida***. Below this, it lists the "Genbank common name: ticks" and the "NCBI BLAST name: mites & ticks". The "Rank" is **order**. The "Genetic code" is [Translation table 1 \(Standard\)](#) and the "Mitochondrial genetic code" is [Translation table 5 \(Invertebrate Mitochondrial\)](#). The "Lineage (full)" is listed as: [cellular organisms](#); [Eukaryota](#); [Opisthokonta](#); [Metazoa](#); [Eumetazoa](#); [Ecdysozoa](#); [Panarthropoda](#); [Arthropoda](#); [Chelicerata](#); [Arachnida](#); [Acari](#).

**Figura 22.** Hierarquia do Lineage do termo *Rhipicephalus microplus*.

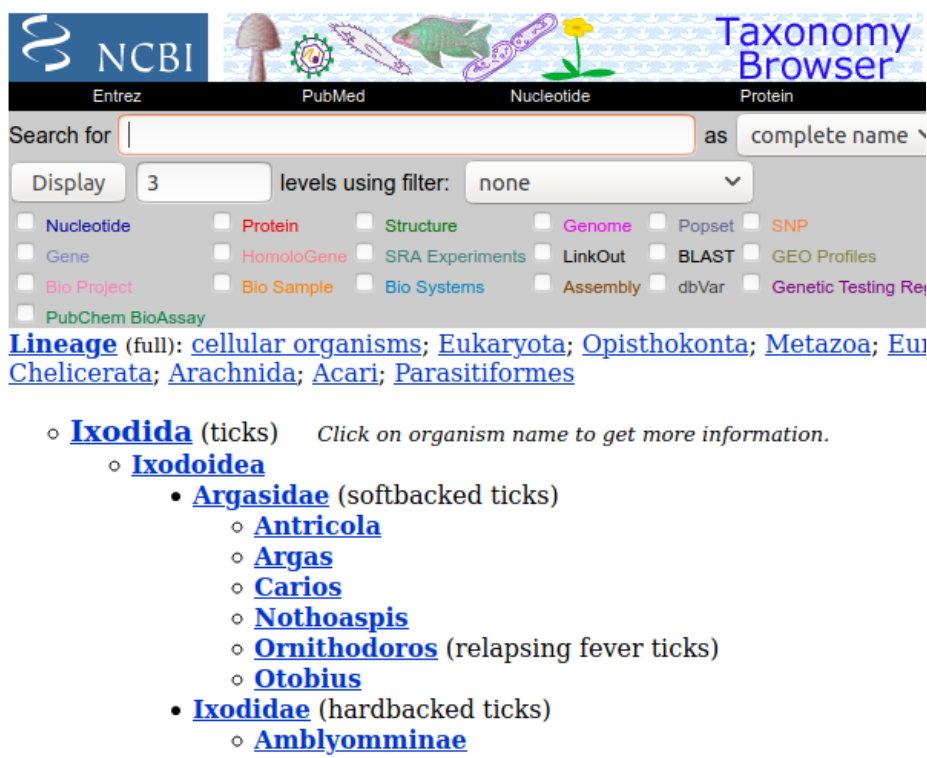


Figura 23. Resultado da escolha do termo Ixodida.

**Passo 4: Obtenção da URL para o termo:** na página de resultado de busca da categoria *Ixodida* apresentada na Figura 23, no frame à direita, em *Taxonomy*, clique no endereço correspondente na coluna *Direct links* para abrir a página da referida categoria. Em seguida, clique em *Ixodida* para abrir a página desta categoria. Copie a URL desta página e cole-a no campo **URL do termo de classificação** do Elemento Classificação, do formulário de catalogação do Repositório.

**Etapa 5: Preenchimento dos campos do elemento Classificação:** no formulário de catalogação do *dataset* no AgDados, preencha os campos conforme o exemplo:

Termo de classificação: Ixodida

Vocabulário controlado da classificação: NCBI Taxonomy

URL do termo de classificação:

<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=6935>

## Referências

CUNHA, M. B. da; CAVALCANTI, C. R. de O. **Dicionário de biblioteconomia e arquivologia**. Brasília, DF: Briquet de Lemos, 2008. 451 p.

ESTADOS UNIDOS. Department of Agriculture. **NAL Agricultural Thesaurus and Glossary**. Disponível em: <<https://agclass.nal.usda.gov/agt.shtml>>. Acesso em: 8 jul. 2020.

FAO. **AGROVOC**. Rome, 2020. Disponível em: <<http://aims.fao.org/vest-registry/vocabularies/agrovoc>>. Acesso em: 8 jul. 2020.

HENNING, P. C.; RIBEIRO, C. J. S.; SALES, L. F.; MOREIRA, J. L. R.; SANTOS, L. O. B. da S. Desmistificando os princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, v. 14, n. 3, p. 175-192, 2019. DOI: 10.22478/ufpb.1981-0695.2019v14n3.46969.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION (Estados Unidos). **Taxonomy**. Bethesda. Disponível em: <<https://www.ncbi.nlm.nih.gov/taxonomy>>. Acesso em: 10 jul. 2020.

VEIGA, V. Gestão de dados de pesquisa FAIR: dando um JUMP em seus dados. In: ENCONTRO DA REDE SUDESTE DE REPOSITÓRIOS INSTITUCIONAIS, 1., 2019, Rio de Janeiro. **Anais...** Rio de Janeiro: UFRJ, 2019. 59 p. Disponível em: <<https://www.arca.fiocruz.br/handle/icit/33343>>. Acesso em: 8 jul. 2020.

## Apêndice B - Catalogação de *Dataset*: Guia Resumido<sup>7</sup>

### Introdução

O processo de catalogação de *datasets* se desenrola no ambiente do software *Dataverse*, utilizando-se um formulário para descrição dos metadados e upload de arquivos de dados. O software *Dataverse* oferece ao usuário a possibilidade de escolher qual conjunto de metadados deseja usar para melhor atender aos interesses e o contexto dos dados a serem descritos. O conjunto de metadados *Citation Metadata*, doravante denominado Metadados de Citação, é default do software *Dataverse*, sendo exportável para o padrão Dublin Core. Por conseguinte, este é o conjunto de metadados adotado no Projeto-piloto do LMB para a catalogação de *datasets*. Outro destacado conjunto de metadados é o *Life Science Metadata*, que após análise pela equipe do Projeto-piloto também foi escolhido para catalogar *datasets*, em razão da sua especificidade temática (ciências da vida), considerada adequada aos projetos de pesquisa do LMB. Desse modo, Metadados de Citação e Metadados Life Science são utilizados concomitantemente, visto que o software *Dataverse* possibilita o uso dos dois conjuntos de metadados no mesmo formulário de catalogação de *datasets*.

Como em quaisquer conjuntos de metadados, o processo de catalogação de *dataset* deve seguir orientações e regras específicas de catalogação<sup>8</sup> para descrever os elementos metadados e respectivos campos e subcampos. O detalhamento desses dois conjuntos de metadados - Metadados de Citação e Metadados *Life Science* -, com elementos, campos e subcampos, é apresentado de forma resumida neste documento, com breve descrição, indicação de exigência de preenchimento e exemplos.

Este Apêndice B é um instrumento sucinto de ajuda rápida ao catalogador no momento da catalogação de *dataset*, no AgDados. Ou seja, apresenta de forma resumida as orientações de catalogação de *dataset* no AgDados, no âmbito do Projeto-piloto de Implantação de Gestão de Dados de Pesquisa do LMB. As orientações completas estão expostas no corpo principal desta publicação.

---

<sup>7</sup> Versão resumida de “Catalogação de *Dataset* no Repositório de Dados da Embrapa: a Experiência do Projeto-piloto de Implantação de Gestão de Dados de Pesquisa no Laboratório Multiusuário de Bioinformática da Embrapa”, documento-âncora desta publicação.

<sup>8</sup> As orientações para catalogação de *dataset* foram elaboradas a partir de recomendações técnicas quanto ao uso das regras básicas do Código de Catalogação Anglo-Americano - AACR2 (Ribeiro, 1995, 2003; Código..., 2004).

**Nota de esclarecimento:** Em relação à exigência de preenchimento dos campos, a equipe do Projeto-piloto decidiu estabelecer mais dois níveis de exigência (fortemente recomendado e recomendado). O acréscimo desses dois níveis objetivou a ampliação da capacidade de recuperação, reúso e integração dos *datasets* com outros conjuntos e repositórios, de acordo com os princípios FAIR<sup>9</sup>. Uma descrição das exigências de preenchimento encontra-se na Tabela 1.

**Tabela 1.** Exigências de preenchimento dos elementos e campos dos metadados.

Exigência de Preenchimento	Descrição
Requerido	Campo deve ser preenchido. Software Dataverse exige o preenchimento.
Fortemente recomendado	Campo de preenchimento fortemente recomendável. Embora não seja exigência do software, seu preenchimento fornecerá informações para análise de indicadores de qualidade dos metadados.
Recomendado	Campo de preenchimento recomendado. Embora não seja exigência do software, o seu preenchimento fornecerá informações para aumentar o nível de qualidade dos metadados.
Opcional	Campo de preenchimento opcional, ficando a critério do catalogador o fornecimento de informações adicionais. Preencher caso pertinente.

<sup>9</sup> FAIR é o acrônimo de *Findable, Accessible, Interoperable, Reusable*, princípios que preconizam que os dados necessitam ser encontráveis, acessíveis, interoperáveis e reutilizáveis (Go Fair, 2016).

## Descrição de Metadados de Citação

Campos	Definição	Exigência de preenchimento	Exemplos
<b>Título</b>	Nome atribuído ao dataset, e pelo qual será formalmente conhecido. Estrutura do título: Tipo de estudo ômico + Amostra estudada + Espécie estudada (nome comum + nome científico).	Requerido	Transcriptoma da glândula salivar do carrapato bovino ( <i>Rhipicephalus</i> (B.) <i>microplus</i> )
Subtítulo	Nome secundário atribuído ao <i>dataset</i> .	Opcional	Expressão gênica na saliva do carrapato bovino.
Título alternativo	Nome alternativo para o dataset; um nome pelo qual conhecido ou referido; um nome abreviado..	Opcional	Genoma do carrapato
<b>URL alternativa</b>	URL alternativa válida onde o <i>dataset</i> pode ser visualizado.	Recomendado	<a href="https://www.frontiersin.org/articles/10.3389/fcimb.2019.00477/full#supplementary-material">https://www.frontiersin.org/articles/10.3389/fcimb.2019.00477/full#supplementary-material</a>
<b>Outro identificador</b>	Outro identificador digital único existente para identificar o <i>dataset</i> .	Não se aplica	Não se aplica
Agência	Nome da agência geradora do identificador digital.	Opcional	NCBI
Identificador	Outro identificador digital persistente que identifica o dataset.	Opcional	PRJNA596777
<b>Autor</b>	Nome de pessoas, instituições ou agências responsáveis pela criação do <i>dataset</i> .	Não se aplica	Não se aplica
Nome	Nome(s) do autor(s) do dataset. Pode(m) ser autor(es) pessoal(is) ou autor corporativo.	Requerido	Pereira Júnior, João Silva, José
Afiliação	Nome da instituição de vínculo do(s) autor(es) pessoal(is).	Recomendado	Embrapa Gado de Corte
Sistema do Identificador	Sistema do identificador digital do autor.	Recomendado	ORCID ResearcherID Scopus ID
Identificador	Identificador digital persistente do(s) autor(es) pessoal(is) do dataset.	Recomendado	<a href="http://orcid.org/0000-0002-1825-0097">http://orcid.org/0000-0002-1825-0097</a>
<b>Contato</b>	Nome de pessoas, instituições ou agências para contato sobre o <i>dataset</i> .	Não se aplica	Não se aplica
Nome	Nome pessoal ou corporativo para contato sobre o <i>dataset</i> .	Fortemente recomendado	José Silva

Afiliação	Nome da instituição à qual a pessoa está vinculada.	Recomendado	Embrapa Gado de Corte
E-mail do contato	Endereço de e-mail do contato, seja pessoal ou corporativo.	Requerido	jose.silva@empresa.com.br
<b>Descrição</b>	Texto sucinto contendo o propósito, a natureza e o escopo do <i>dataset</i> .	Não se aplica	Não se aplica
Texto	Texto sucinto contendo o propósito, a natureza e o escopo do <i>datasets</i> .	Requerido	Conjunto de dados réplica 1 da glândula salivar Holandês, dados brutos formato fastq.  Sequence name; Sequence description; Sequence length; Blast hits count; Blast min e-value; Blast similarity mean; Blast hit description (HSP); Blast hit taxonomy name; Enzyme code; Enzyme name; InterPro accession.
Data	Data em que a descrição do <i>dataset</i> foi realizada.	Opcional	2020-06-08
<b>Assunto</b>	Áreas de domínio específicas e relevantes para o <i>dataset</i> .	Requerido	Agricultural Sciences Medicine, Health and Life Sciences
<b>Palavras-chave</b>	Palavras que expressam o conteúdo-chave do <i>dataset</i> , sendo capazes de representar aspectos relevantes para a compreensão do assunto.	Não se aplica	Não se aplica
Termo-chave	Termo-chave	Fortemente recomendado	Carrapato Host-parasite relationship Nucleotide sequences Ticks Transcriptome
Vocabulário-controlado	Vocabulário controlado utilizado para a atribuição de termo-chave	Fortemente recomendado	AGROVOC Thesaurus NAL Thesaurus NCBI Taxonomy
URL do termo-chave	URL do termo-chave onde o mesmo possa ser acessado.	Fortemente recomendado	URL da palavra-chave 'nucleotide sequence' (AGROVOC Thesaurus): <a href="http://aims.fao.org/aos/agrovoc/c_27583">http://aims.fao.org/aos/agrovoc/c_27583</a>  URL da palavra-chave 'nucleotide sequences' (NAL Thesaurus): <a href="https://agclass.nal.usda.gov/mtw/dk.exe?k=default&amp;l=60&amp;w=15979&amp;s=5&amp;t=2">https://agclass.nal.usda.gov/mtw/dk.exe?k=default&amp;l=60&amp;w=15979&amp;s=5&amp;t=2</a>
<b>Classificação</b>	Refere-se à classificação do assunto do <i>dataset</i> , por meio da atribuição de uma ou mais categorias, sob as quais podem ser agrupados <i>datasets</i> de natureza semelhante.	Não se aplica	Não se aplica

Termo de classificação	Indicação da classificação de assunto (categoria).	Fortemente recomendado	L Animal Science and Animal Products (NAL Thesaurus) Phenomena (AGROVOC Thesaurus) Ixodida (NCBI Taxonomy)
Vocabulário controlado	Vocabulário controlado utilizado para a classificação de assunto.	Fortemente recomendado	AGROVOC Thesaurus NAL Thesaurus NCBI Taxonomy
URL do termo	URL do termo de classificação.	Fortemente recomendado	URL do termo de classificação 'phenomena' (AGROVOC Thesaurus): <a href="http://aims.fao.org/aos/agrovoc/c_330704">http://aims.fao.org/aos/agrovoc/c_330704</a>  URL do termo de classificação 'L Animal Science and Animal Products' (NAL Thesaurus): <a href="https://agclass.nal.usda.gov/mtw/dk.exe?k=default&amp;l=60&amp;s=cid&amp;w=100">https://agclass.nal.usda.gov/mtw/dk.exe?k=default&amp;l=60&amp;s=cid&amp;w=100</a>  URL do termo de classificação 'Ixodida' (NCBI Taxonomy): <a href="https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Tree&amp;id=6935&amp;lvl=3&amp;p=nucleocore&amp;keep=1&amp;srchmode=1&amp;unlock">https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Tree&amp;id=6935&amp;lvl=3&amp;p=nucleocore&amp;keep=1&amp;srchmode=1&amp;unlock</a>
<b>Publicação relacionada</b>	Publicações que foram produzidas a partir de dados do <i>dataset</i> .	Não se aplica	Não se aplica
Citação	Referência completa da publicação relacionada ao <i>dataset</i> .	Recomendado	GIACHETTO, P. F. Transcriptoma do carrapato dos bovinos. In: GARCIA, M. V.; KOLLER, W. W. (Ed.). <b>Carrapatos na cadeia produtiva de bovinos</b> . Brasília, DF: Embrapa, 2019. cap. 16. p. 205-222.
Tipo de identificador	Tipo de identificador digital usado pela publicação relacionada.	Recomendado	DOI
Número do identificador	Número do identificador digital único da publicação relacionada.	Recomendado	<a href="https://doi.org/10.3389/fcimb.2019.00477">https://doi.org/10.3389/fcimb.2019.00477</a>
URL da publicação	URL da página web onde se encontra a publicação.	Opcional	<a href="https://www.frontiersin.org/journals/cellular-and-infection-microbiology">https://www.frontiersin.org/journals/cellular-and-infection-microbiology</a>
<b>Notas</b>	Informações adicionais sobre o <i>dataset</i> .	Opcional	Réplica de dados de glândula salivar do carrapato bovino infestando gado Holandês. Ensaio envolve dados de gado Nelore e mestiço Nelore x Holandês.
<b>Idioma</b>	Idioma do <i>dataset</i>	Opcional	Inglês Inglês Português



			Inglês Português Espanhol
<b>Produtor</b>	Pessoa ou instituição com responsabilidade financeira ou administrativa sobre o <i>dataset</i> .	Não se aplica	Não se aplica
Nome do produtor	Nome pessoal ou corporativo do produtor do <i>dataset</i> .	Recomendado	José Silva Empresa Brasileira de Pesquisa Agropecuária
Afiliação	Nome da instituição de vínculo do produtor.	Recomendado	Embrapa LMB
Abreviatura	Abreviatura do nome do produtor caso seja um nome corporativo.	Opcional	Embrapa Gado de Corte Empresa Brasileira de Pesquisa Agropecuária
URL do produtor	URL válida da página na web do produtor do <i>dataset</i> .	Opcional	URL Empresa Brasileira de Pesquisa Agropecuária: <a href="https://www.embrapa.br/">https://www.embrapa.br/</a>  URL do Laboratório Multiusuário de Bioinformática: <a href="https://www.embrapa.br/laboratorio-de-bioinformatica">https://www.embrapa.br/laboratorio-de-bioinformatica</a>
URL do logotipo do produtor	URL válida da página web onde o logotipo do produtor do dataset pode ser visualizado.	Opcional	<a href="https://www.embrapa.br/image/layout_set_logo?img_id=1343656&amp;t=1591812010844">https://www.embrapa.br/image/layout_set_logo?img_id=1343656&amp;t=1591812010844</a>
<b>Data da produção</b>	Data em que o <i>dataset</i> foi produzido.	Fortemente recomendado	2015-05-01
<b>Local de produção</b>	Local onde o <i>dataset</i> e quaisquer outros materiais relacionados foram produzidos.	Recomendado	Laboratório de Biotecnologia Animal Centro de Genômica Funcional
<b>Colaborador</b>	Pessoa ou instituição responsável pela coleta, gestão ou qualquer outra contribuição relevante para o desenvolvimento do <i>dataset</i> .	Não se aplica	Não se aplica
Tipo	Tipo de colaborador do <i>dataset</i> .	Opcional	Coletor de dados Curador de dados Líder do projeto Pesquisador
Nome	Nome pessoal ou corporativo do colaborador do <i>dataset</i> .	Fortemente recomendado	Nome pessoal de colaborador: Paula Regina Kuser Falcão  Nome de corporativo (instituição) de colaborador: Laboratório de Biotecnologia Animal Centro de Genômica Funcional
<b>Financiamento</b>	Informações sobre instituição, agência de fomento, organização e projetos responsável pela concessão de suporte e/ou apoio financeiro para a produção.	Não se aplica	Não se aplica

Nome do órgão	Nome da instituição, agência de fomento ou organização que concedeu apoio financeiro para a produção do <i>dataset</i> .	Recomendado	Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) Empresa Brasileira de Pesquisa Agropecuária (Embrapa) Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp)
Número do contrato	Número do contrato de concessão de financiamento	Recomendado	01.11.07.002.03.00 (projeto SEG, da Embrapa)
<b>Distribuidor</b>	Pessoa ou instituição designada pelo autor ou produtor do <i>dataset</i> para gerar cópias do trabalho, incluindo quaisquer edições ou revisões necessárias.	Não se aplica	Não se aplica
Nome	Nome pessoal ou corporativo do distribuidor do <i>dataset</i> .	Fortemente recomendado	Nome pessoal: José Silva Nome de corporativo Empresa Brasileira de Pesquisa Agropecuária
Afiliação	Nome da instituição à qual o distribuidor do <i>dataset</i> está vinculado.	Recomendado	Embrapa
Abreviatura	Abreviatura do nome do produtor caso seja um nome corporativo.	Opcional	Embrapa LMB
URL do distribuidor	URL válida da página na web do distribuidor do <i>dataset</i> .	Opcional	URL da Empresa Brasileira de Pesquisa Agropecuária: <a href="https://www.embrapa.br/">https://www.embrapa.br/</a>  URL do Laboratório Multiusuário de Bioinformática: <a href="https://www.embrapa.br/laboratorio-de-bioinformatica">https://www.embrapa.br/laboratorio-de-bioinformatica</a>
URL do logotipo	URL válida da página web onde o logotipo do distribuidor do <i>dataset</i> pode ser visualizado.	Opcional	URL do logotipo da Embrapa: <a href="https://www.embrapa.br/image/layout_set_logo?img_id=1343656&amp;t=1591812010844">https://www.embrapa.br/image/layout_set_logo?img_id=1343656&amp;t=1591812010844</a> URL do logotipo do Laboratório Multiusuário de Bioinformática: <a href="https://www.dataverse-h.cnptia.embrapa.br/logos/8/layout_set_logo.png">https://www.dataverse-h.cnptia.embrapa.br/logos/8/layout_set_logo.png</a>
<b>Data de distribuição</b>	Data em que o <i>dataset</i> foi disponibilizado para distribuição.	Recomendado	2020-01-15
<b>Depositante</b>	Nome pessoal ou corporativo do depositante do <i>dataset</i> .	Requerido	Nome pessoal: José Silva Nome corporativo: Empresa Brasileira de Pesquisa Agropecuária
<b>Data do depósito</b>	Data em que o <i>dataset</i> foi depositado no repositório.	Recomendado	2020-06-11

<b>Período de cobertura dos dados</b>	Período de tempo ao qual os dados se referem ou período de tempo abrangido pelos dados.	Não se aplica	Não se aplica
Data de início	Data de início do período de cobertura dos dados.	Recomendado	2015-05-01
Data de Término	Data de término do período da cobertura dos dados.	Recomendado	2015-06-22
<b>Período de coleta dos dados</b>	Período em que os dados foram coletados.	Não se aplica	Não se aplica
Data de início	Data de início do período da coleta dos dados.	Recomendado	2015-05-01
Data de Término	Data de término do período da coleta de dados.	Recomendado	2015-06-22
<b>Tipos de dados</b>	Tipo de dados incluídos no <i>dataset</i> .	Recomendado	Código-fonte Dados agregados Dados de pesquisa Dados experimentais Dados biológicos Dados genômicos Dados de sequência RNA
<b>Série</b>	Informação sobre a série do <i>dataset</i> .	Não se aplica	Não se aplica
Nome	Nome de série ao qual o <i>dataset</i> está vinculado	Opcional	Portfólio de Sanidade Animal Manejo de Pragas e Doenças
Informação	Informação sucinta que descreve as características da série.	Opcional	A Série Portfólio Sanidade Animal e Manejo de Pragas e Doenças abarca o conjunto de <i>dataset</i> produzidos no âmbito do Projeto RmVac.
<b>Software</b>	Informações sobre o software usado para gerar o <i>dataset</i> .	Não se aplica	Não se aplica
Nome	Nome do software utilizado para gerar o <i>dataset</i> .	Recomendado	SAS Scripts Software R Trinity
Versão	Versão do software.	Recomendado	Versão 2.8.4 Versão 0.01
<b>Material relacionado</b>	Outro material relacionado ao <i>dataset</i> que está sendo descrito.	Opcional	Projetos anteriores <i>Datasets</i> anteriores
<b><i>Datasets</i> relacionados</b>	Quaisquer <i>datasets</i> relacionados ao <i>dataset</i> que está sendo descrito, tais como <i>datasets</i> provenientes de pesquisas anteriores sobre esse assunto.	Opcional	Projetos anteriores
<b>Outras referências</b>	Quaisquer outras referências que podem ter servido de material de apoio ao <i>dataset</i> .	Opcional	Acesso e descoberta de conhecimento produzido por pesquisadores de instituições parceiras do Projeto RmVac.

<b>Fontes de dados</b>	Referem-se a materiais que serviram de fontes para o <i>dataset</i> , como: lista de livros, artigos, publicações em série ou arquivos de dados legíveis por máquina.	Opcional	CSORDAS, B. G.; CUNHA, R. C.; GARCIA, M. V.; SILVA, S. S. da; LEITE, F. L.; ANDREOTTI, R. Molecular characterization of the recombinant protein RmLTIBmCG-LTB: protective immunity against <i>Rhipicephalus (Boophilus) microplus</i> . <b>PLoS One</b> , v. 13, n. 2, p. 1-14, 2018. DOI: 10.1371/journal.pone.0191596
<b>Origem das fontes de dados</b>	Para materiais históricos, informar a origem das fontes e as regras seguidas no estabelecimento das fontes do <i>dataset</i> .	Opcional	Literatura que retrata o estado da arte.
<b>Características das fontes observadas</b>	Características e material de origem do <i>dataset</i> .	Opcional	Não há
<b>Documentação e acesso a fontes de dados</b>	Nível de documentação do material de origem. Informação sobre quaisquer restrições de acesso aos dados de origem.	Opcional	Não há

## Descrição de Metadados *Life Science*

Campos	Definição	Exigência de preenchimento	Exemplos
<b>Tipo de design</b>	Definição geral do desenho experimental.	Opcional	Case control; Cross sectional; Cohort study; Nested case control design; Not specified; Parallel group design; Perturbation design; Randomized controlled trial; Technological design.
<b>Tipo de fator</b>	Fatores utilizados no <i>dataset</i> . Fator experimental; estresse hídrico; estágio de desenvolvimento; tipo de célula.	Opcional	Age; Biomarkers; Cell surface markers; Cell type/cell line; Developmental stage; Genetic characteristics; Treatment type etc.
<b>Organismo</b>	Nome taxonômico do organismo usado no conjunto de dados ou de onde o material biológico se originou	Opcional	<i>Bos taurus</i> <i>Escherichia coli</i> <i>Rhipicephalus microplus</i> Carrapato do boi
<b>Outro organismo</b>	Não se aplica.	Não se aplica.	Não se aplica.
<b>Tipo de medida</b>	Um termo para qualificar o que está sendo medido.	Opcional	Cell counting; DNA methylation profiling; Genome sequencing; Protein=RNA binding (RIP-Seq) etc.
<b>Outro tipo de medida</b>	Não se aplica.	Não se aplica.	Não se aplica.
<b>Tipo de tecnologia</b>	Um termo usado para identificar a tecnologia usada para coletar os dados.	Opcional	DNA microarray; Mass spectrometry; Nucleotide sequencing; Protein microarray etc.
<b>Plataforma tecnológica</b>	O fabricante e o nome da plataforma usada no experimento.	Opcional	Illumina HiSeq2500
<b>Tipo de célula</b>	Nome da linhagem de células das quais as amostras derivam.	Opcional	Larvas; Ninfas; Ovários; Intestinos; Glândula salivar etc.

## Referências

CÓDIGO de catalogação anglo-americano. 2. ed., rev. 2002. São Paulo: FEBAB: Imprensa Oficial do Estado de São Paulo, 2004. 1 v. Paginação irregular.

GO FAIR. **FAIR principles**. [2016]. Disponível em: <<https://www.go-fair.org/fair-principles/>>. Acesso em: 27 ago. 2020.

RIBEIRO, A. M. de C. M. **AACR2**: Anglo-American cataloguing rules, 2nd edition: descrição e pontos de acesso. Brasília, DF: Ed. do Autor, 1995. 577 p.

RIBEIRO, A. M. de C. M. **Catálogo de recursos bibliográficos pela AACR2 2002**: Anglo-American Cataloguing Rules, 2nd edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular.

## Apêndice C – Revisão na Catalogação de *Dataset*

### Introdução

O Apêndice C traz orientações para a execução da atividade de revisão na catalogação de *datasets*, no Repositório de Dados da Embrapa (AgDados). Essas orientações foram elaboradas no âmbito do Projeto-piloto de Implantação de Gestão de Dados de Pesquisa no Laboratório Multiusuário de Bioinformática da Embrapa (LMB), com o objetivo de apoiar o trabalho de revisores (bibliotecários e cientistas da informação) na análise de qualidade na catalogação de metadados e *datasets*.

Neste apêndice serão abordados os seguintes tópicos:

- Catalogação de *datasets* no AgDados.
- Por que revisar a catalogação de *datasets* no AgDados?
- Controle de qualidade e garantia de qualidade.
- Revisão na catalogação de *datasets* no AgDados.
- Atividades de revisão na catalogação de *datasets*.

### Catalogação de *dataset* no AgDados

A catalogação de *datasets* é uma atividade essencialmente descritiva, inerente à etapa 'Descrição', presente nos principais modelos de ciclo de vida de dados, como pode ser observado em estudos realizados por Ball (2012) e Sinaeepourfar et al. (2015). A catalogação é uma atividade que, além de descrever, visa conferir qualidade tanto aos metadados quanto aos dados (*datasets*) em si. A catalogação é uma atividade capaz de gerar impactos em outros processos dela decorrentes, por exemplo, na indexação, na busca e na recuperação de informações, para os quais os vocabulários controlados para tratamento de assuntos, tornam-se imprescindíveis e de uso obrigatório na atribuição de palavra-chave e classificação (ver Apêndice A).

No escopo do Projeto-piloto de Implementação de Gestão de Dados de Pesquisa no Laboratório Multiusuário de Bioinformática da Embrapa (LMB), a equipe do projeto escolheu o processo de catalogação para ser o objeto de análise e melhorias, visando sua eficiência e eficácia. Portanto, este é um processo que envolve todas as atividades relativas à representação descritiva e temática de *datasets*.

O processo de catalogação de *datasets* que se desenrola no ambiente do software Dataverse utiliza um formulário para descrição dos metadados e upload de arquivos de dados. O conjunto de Metadados de Citação, default do software Dataverse, é utilizado para a

catalogação de *datasets*. Por conseguinte, este é o conjunto de metadados adotado no Projeto-piloto do LMB para a catalogação de *datasets*. Outro destacado conjunto de metadados é o Life Science Metadata, que também foi escolhido para catalogar *datasets*. Desse modo, Metadados de Citação e Metadados Life Science são utilizados concomitantemente, dada a possibilidade de o software Dataverse combinar o uso dos dois conjuntos de metadados no mesmo formulário de catalogação de *datasets*.

Para a representação descritiva são adotadas as regras mínimas da AACR2 (Ribeiro, 1995, 2003; Código..., 2004); e para a representação temática são adotados os vocabulários controlados AGROVOC (FAO, 2020), NAL Thesaurus (Estados Unidos, 2020) e NCBI Taxonomy (National Center for Biotechnology Information, 2020).

## Por que revisar a catalogação de *dataset* no AgDados?

A busca de qualidade em catalogação deve ser assegurada por meio da adoção de padrões mínimos e universais, sobretudo, de descrição e indexação, os quais contribuem para a coerência e a consistência do registro no repositório e para a sua efetiva recuperação. (Souza; Mostafa, 1999 citado por Branco et al., 2014). De acordo com Graham (1990), completeza e precisão do registro são os elementos que caracterizam a qualidade da catalogação. A completeza expressa o quanto os metadados representam o documento descrito. A precisão diz respeito ao corretismo das informações fornecidas na descrição do recurso digital (Morris; Wool, 1999).

A revisão na catalogação de *datasets* é uma atividade que contribui para que os metadados se tornem aderentes aos princípios FAIR<sup>10</sup>, e, consequentemente, possam usufruir dos benefícios de serem encontráveis, acessíveis, interoperáveis e reutilizáveis tanto para humanos como para máquinas.

## Controle de qualidade e garantia de qualidade

Controle de qualidade diz respeito à avaliação de um produto, processo, atividade ou serviço. Por sua vez, a garantia de qualidade visa assegurar que determinado produto, serviço ou processo atinja níveis mínimos de qualidade. Ou seja, a garantia da qualidade está empenhada em fazer com que produto, processo, atividade ou serviço atinja de forma eficiente os objetivos traçados. (Branco et al., 2014).

No tocante a determinado processo, por exemplo, o controle de qualidade preocupa-se em assegurar o atendimento aos requisitos técnicos específicos, como também ser

---

<sup>10</sup> FAIR é o acrônimo de *Findable, Accessible, Interoperable, Reusable*, princípios que preconizam que os dados necessitam ser encontráveis, acessíveis, interoperáveis e reutilizáveis (Go Fair, 2016).



confiável e satisfatório para os usuários. Desse modo, o processo de controle de qualidade envolve, necessariamente, a atividade de análise e identificação de atividades e de processos que possam estar em descumprimento de normas específicas estabelecidas para a execução. (Branco et al., 2014).

Após a identificação do processo ou atividade para o qual se pretende exercer o controle sobre sua qualidade, Branco et al. (2014, p. 7) recomendam os seguintes passos: a) definir requisitos e critérios para avaliar a efetividade do controle de qualidade; b) estabelecer diretrizes para o exercício do controle de qualidade; c) monitorar a manutenção e o desenvolvimento da atividade de controle de qualidade; d) qualificar os sujeitos envolvidos no processo de controle de qualidade; e f) capacitar profissionais para a melhoria de produto, serviço, atividade e processo.

A garantia de qualidade visa assegurar que determinado produto, serviço ou processo atinja níveis mínimos de qualidade. No tocante ao processo de catalogação de *datasets*, a descrição padronizada de metadados é fator crítico de garantia de qualidade.

## Revisão de catalogação de *dataset* no AgDados

A catalogação de *dataset* no Repositório de Dados da Embrapa (AgDados) é realizada no ambiente do software Dataverse. Para tanto, são adotados padrões internacionais para a representação descritiva e vocabulários controlados para a representação temática, mencionados anteriormente.

A busca de qualidade na catalogação de *dataset* deve ser assegurada por medidas de controle voltadas ao atendimento a padrões e regras mínimas de catalogação, no tocante ao preenchimento de elementos, campos e subcampos do formulário de catalogação no software Dataverse. Para tanto, duas medidas são necessárias: a) observância ao documento orientador 'Catalogação de *datasets* no Repositório de Dados da Embrapa (AgDados); b) estabelecimento dos elementos de controle de qualidade na catalogação <sup>11</sup>.

Para o estabelecimento dos elementos de controle de qualidade, baseando-se no estudo realizado por Branco et al. (2014), foram escolhidos os principais elementos necessários para o controle de qualidade na catalogação de *datasets* no AgDados, mostrados na Tabela 1.

---

<sup>11</sup> O documento orientador mencionado anteriormente corresponde ao conteúdo na íntegra dos tópicos 1, 2, 3 e 4, bem como os Apêndices A e B, desta publicação.

**Tabela 1.** Elementos de controle de qualidade para a revisão em catalogação de *datasets*

Elementos de controle
Completeza das informações registradas sobre o <i>dataset</i>
Precisão da descrição quanto: ortografia, pontuação, descrição textual, transcrições de URI, URL, DOI, ORCID etc.
Clareza e integridade das informações registradas sobre o <i>dataset</i> .
Fidelidade aos padrões e normas adotados para a catalogação de <i>dataset</i> .
Confiabilidade das informações registradas sobre o <i>dataset</i> .
Coerência e consistência dos dados e informações registradas sobre o <i>dataset</i> .

Como pode ser inferido na Tabela 1, faz-se relevante enfatizar a necessidade de adoção de padrões e orientações técnicas prescritas para a catalogação no AgDados, para que sejam obtidos níveis de qualidade aceitáveis no que diz respeito à qualidade das informações sobre os *datasets* catalogados.

### Atividades de revisão na catalogação de *dataset*

Dada a natureza da atividade, o controle de qualidade em catalogação de *datasets* deve ser executado por profissionais de Biblioteconomia e/ou Ciência da Informação. A revisão envolve a análise acurada do preenchimento de cada elemento, campo e subcampo do formulário de catalogação no software Dataverse.

A tarefa de revisão e análise será guiada pelos elementos de controle de qualidade estabelecidos na Tabela 1, devendo o revisor observar as etapas mostradas na Tabela 2, a seguir.

**Tabela 2.** Etapas de controle de qualidade na catalogação de *datasets*.

Etapas de controle
1) Examina todas as informações registradas sobre o <i>dataset</i> .
2) Descreve todos os erros e/ou inconsistências encontrados.
3) Notifica o catalogador
4) Solicita correção de erros e/ou inconsistências identificados.
5) Mantém registro de correções efetuadas.
6) Encaminha relatório para o catalogador.

A atividade de revisão da catalogação de *datasets* deverá ser realizada em conformidade com as matrizes de revisão apresentadas nas Tabelas 3 e 4, abaixo, elaboradas

com base na Tabela 1, acima. Os apontamentos para a correção da catalogação deverão ser registrados em formulário próprio (Anexo 1 - Relatório de apontamentos para a correção da catalogação de *datasets*), indicando-se os erros e/ou inconsistências e as orientações para a correção, além dos impactos na qualidade (Tabela 1).

**Tabela 3.** Matriz de atividades de revisão na catalogação de *dataset* no AgDados - Metadados de Citação.

Elemento (campo/subcampo)	Descrição do erro e/ou inconsistência	Orientação para correção	Elemento(s) de impacto na qualidade	Documento orientador
<b>Título</b>	Ausência de clareza e integridade na atribuição do título.	Atribuir título observando: Tipo de estudo; Amostra estudada; Espécie estudada.	Completeza Clareza Integridade	Catálogo de <i>dataset</i> no AgDados
Subtítulo	Ausência de informação do subtítulo.	Informar subtítulo.	Completeza	Catálogo de <i>dataset</i> no AgDados
Título alternativo	Ausência de informação do título alternativo.	Informar título alternativo.	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>URL alternativa</b>	Ausência da URL alternativa.	Informar URL alternativa válida.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Outro identificador</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Agência	Ausência do nome da agência do outro identificador.	Informar nome da agência do outro identificador.	Completeza	Catálogo de <i>dataset</i> no AgDados
Identificador	Ausência de identificador digital. Identificador digital não localizado.	Informar identificador digital válido.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Autor</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Nome	Ausência de padronização na descrição do nome.	Seguir regras de catalogação.	Fidelidade a padrões	Catálogo de <i>dataset</i> no AgDados
Afiliação	Ausência de padrão na descrição de nome corporativo.	Seguir regras de catalogação.	Fidelidade a padrões	Catálogo de <i>dataset</i> no AgDados
Sistema identificador	Ausência de indicação de sistema identificador.	Informar nome do sistema identificador.	Completeza	Catálogo de <i>dataset</i> no AgDados
Identificador	Ausência de identificador persistente válido.	Informar nome identificador persistente.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Contato</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica

Nome	Ausência de padronização na descrição do nome.	Seguir regras de catalogação.	Fidelidade a padrões	Catalogação de <i>dataset</i> no AgDados
Afiliação	Ausência de padrão na descrição de nome corporativo.	Seguir regras de catalogação.	Fidelidade a padrões	Catalogação de <i>dataset</i> no AgDados
E-mail do contato	E-mail incompleto.	Informar e-mail válido.	Precisão	Catalogação de <i>dataset</i> no AgDados
<b>Descrição</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Texto	Ausência de clareza e coerência textual no campo Texto Ausência de precisão na descrição quanto à ortografia, pontuação e transcrição textual na descrição.	Revisar a redação do campo Texto, observando clareza, coerência e precisão quanto à ortografia, pontuação e transcrição textual.	Clareza e integridade. Coerência e consistência	Catalogação de <i>dataset</i> no AgDados
Data	Ausência de informação da Data de descrição.	Informar Data de descrição.	Completeza	Catalogação de <i>dataset</i> no AgDados
<b>Assunto</b>	Ausência de área(s) de domínio representativa do assunto.	Adicionar outra(s) área(s) de domínio.	Completeza	Catalogação de <i>dataset</i> no AgDados
<b>Palavras-chave</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Termo-chave	Ausência (falta) de termos-chave. Termo-chave selecionado não pertence ao escolhido.	Inserir mais termos-chave Corrigir transcrição da termos-chave	Completeza Confiabilidade da informação Coerência e consistência da informação	Catalogação de <i>dataset</i> no AgDados
Vocabulário controlado	Ausência (falta) de indicação de nome(s) do(s) vocabulário(s) controlado(s) usado(s) para o(s) termo(s)-chave.	Inserir o nome do vocabulário(s) controlado(s) para usado para indicar o(s) termo(s)-chave.	Completeza da informação	Catalogação de <i>dataset</i> no AgDados
URL da palavra-chave	Ausência da URL da palavra-chave. URL da palavra-chave não localizada.	Informar URL válida da palavra-chave	Completeza Precisão	Catalogação de <i>dataset</i> no AgDados
<b>Classificação</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Termo de classificação	Ausência de termos de classificação. Termo de classificação escolhido não	Adicionar mais termos de classificação Eliminar termo de classificação não	Completeza Confiabilidade da informação	Catalogação de <i>dataset</i> no AgDados

	corresponde ao assunto do <i>dataset</i> .	correspondente ao tema do <i>dataset</i> .	Coerência e consistência da informação	
Vocabulário controlado da classificação	Ausência de nome(s) de vocabulário(s) controlado(s).	Informar nome(s) de vocabulário(s) controlado(s).	Completeza Precisão Fidelidade a padrões	Catálogo de <i>dataset</i> no AgDados
URL do termo de classificação	Ausência da URL de termo(s) de classificação. URL do termo de classificação não válida.	Adicionar URL de termo(s) de classificação válida Informar URL do termo válida.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Publicação relacionada</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Citação	Ausência de padronização na referência da citação.	Seguir regras de referência bibliográfica.	Fidelidade a padrões	Manual de referência para recursos da informação da Embrapa
Tipo de identificador	Ausência de indicação de tipo de identificador.	Informar tipo de identificador.	Completeza	Catálogo de <i>dataset</i> no AgDados
Número do identificador	Ausência de número de identificador válido.	Informar nome identificador válido.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
URL da publicação	Ausência da URL da publicação.	Informar URL válida da publicação.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Notas</b>	Ausência de clareza e coerência textual da nota. Ausência de precisão, quanto à ortografia, pontuação no texto da nota.	Revisar a redação de nota, observando clareza, coerência e precisão textual quanto à ortografia, pontuação.	Clareza e integridade Coerência e consistência	Catálogo de <i>dataset</i> no AgDados
<b>Idioma</b>	Ausência de informação sobre o idioma <i>dataset</i> .	Informar o idioma do <i>dataset</i> .	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Produtor</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Nome do produtor	Ausência de padronização na indicação do nome do produtor.	Seguir regras de catalogação.	Fidelidade a padrões	Catálogo de <i>dataset</i> no AgDados
Afiliação do produtor	Ausência de padrão na descrição de nome corporativo.	Seguir regras de catalogação.	Fidelidade a padrões	Catálogo de <i>dataset</i> no AgDados
Abreviatura do nome do produtor	Ausência de abreviatura do nome do produtor.	Informar a abreviatura do nome do produtor.	Fidelidade a padrões	Catálogo de <i>dataset</i> no AgDados

URL do produtor	Ausência da URL do produtor.	Informar URL válida do produtor.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
URL do logotipo do produtor	Ausência da URL do logo do produtor.	Informar URL válida do logo do produtor.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Data de produção</b>	Ausência de data de produção.	Informar data de produção.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Local de produção</b>	Ausência de local de produção	Informar local de produção.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Colaborador</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Tipo de colaborador	Ausência de informação sobre o tipo de colaborador.	Selecionar tipo de colaborador.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
Nome do colaborador	Ausência de padronização na indicação do nome do colaborador.	Seguir regras de catalogação.	Fidelidade a padrões	Catálogo de <i>dataset</i> no AgDados
<b>Financiamento</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Nome do órgão financiador	Ausência do nome do órgão financiador	Indicar nome do órgão financiador.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
Número do contrato	Ausência do número do contrato de financiamento.	Informar número de contrato de financiamento.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Distribuidor</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Nome do distribuidor	Ausência do nome do distribuidor do <i>dataset</i> . Ausência de padronização na indicação do nome do distribuidor.	Indicar nome do distribuidor. Seguir regras de catalogação na indicação do nome do distribuidor.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
Afiliação do distribuidor	Ausência de padrão na descrição do nome do distribuidor.	Seguir regras de catalogação.	Fidelidade a padrões	Catálogo de <i>dataset</i> no AgDados
Abreviatura do nome do distribuidor	Ausência de abreviatura do nome do distribuidor.	Informar a abreviatura do nome do distribuidor.	Fidelidade a padrões	Catálogo de <i>dataset</i> no AgDados
URL do distribuidor	Ausência da URL do distribuidor.	Informar URL válida do distribuidor.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
URL do logotipo do distribuidor	Ausência da URL do logo do distribuidor.	Informar URL válida do logo do distribuidor.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Data de distribuição</b>	Ausência de data de distribuição.	Informar data de distribuição.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados

<b>Depositante</b>	Ausência de padronização na indicação do nome do depositante.	Seguir regras de catalogação na indicação do nome do depositante.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Data do depósito</b>	Ausência de data de depósito.	Informar data de depósito.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Período da cobertura de dados</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Data de início	Ausência de data de início da cobertura dos dados.	Informar data de início da cobertura dos dados.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
Data de término	Ausência de data de término da cobertura dos dados.	Informar data de término da cobertura dos dados.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Período de coleta de dados</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Data de início	Ausência de data de início da coleta dos dados.	Informar data de início da coleta dos dados.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
Data de término	Ausência de data de término da coleta dos dados.	Informar data de término da coleta dos dados.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Tipo de dados</b>	Ausência de informação do(s) tipo(s) de dados do <i>dataset</i> .	Informar tipo(s) de dados do <i>dataset</i> .	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Série</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Nome da série	Ausência do nome da série do <i>dataset</i> .	Informar nome da série, caso exista.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
Informação sobre série	Ausência de informações sobre a série do <i>dataset</i> . Ausência de clareza e coerência textual na informação sobre a série.	Fornecer informações sucintas sobre a série, caso exista. Revisar redação da informação sobre a série.	Clareza e integridade Coerência e consistência	Catálogo de <i>dataset</i> no AgDados
<b>Software</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Nome de software	Ausência do nome do software utilizado.	Indicar o nome do software	Completeza Confiabilidade das informações	Catálogo de <i>dataset</i> no AgDados
Versão do software	Ausência da informação sobre a versão do software utilizado.	Indicar o nome do software.	Completeza Confiabilidade das informações	Catálogo de <i>dataset</i> no AgDados

<b>Material relacionado</b>	Ausência de informação sobre material relacionado.	Informar material utilizado relacionado ao <i>dataset</i> , caso exista.	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Datasets relacionados</b>	Ausência de informação de <i>datasets</i> relacionados.	Informar <i>dataset</i> relacionados, caso existam.	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Outras referências</b>	Ausência de informação sobre outras referências ao <i>dataset</i> .	Informar outras referências relacionadas ao <i>dataset</i> , caso existam.	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Fontes de dados</b>	Ausência de informação sobre fontes de dados do <i>dataset</i> .	Informar fontes de dados do <i>dataset</i> , caso existam.	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Origem das fontes de dados</b>	Ausência de informação sobre origem das fontes dos dados do <i>datasets</i> .	Informar origem das fontes de dados do <i>dataset</i> , caso exista.	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Características das fontes de dados observadas</b>	Ausência de informação sobre características das fontes de dados observadas.	Informar sobre características das fontes de dados observadas, caso existam.	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Documentação e acesso a fontes de dados</b>	Ausência de informação sobre documentação e acesso a fontes de dados do <i>dataset</i> .	Informar sobre documentação e acesso a fontes de dados do <i>dataset</i> , caso exista.	Completeza	Catálogo de <i>dataset</i> no AgDados

**Tabela 4.** Matriz de atividades de revisão na catalogação de *dataset* no AgDados - Metadados *Life Science*.

Elemento (campo)	Descrição do erro e/ou inconsistência	Orientação para correção	Elemento(s) de impacto na qualidade	Documento orientador
<b>Tipo de design</b>	Ausência de informação	Selecionar tipo de design	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Tipo de fator</b>	Ausência de informação	Selecionar tipo de fator	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Organismo</b>	Ausência de informação	Selecionar organismo	Completeza	Catálogo de <i>dataset</i> no AgDados



<b>Outro organismo</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
<b>Tipo de medida</b>	Ausência de informação	Selecionar tipo de medida	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Outro tipo de medida</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
<b>Tipo de tecnologia</b>	Ausência de informação	Selecionar tipo de tecnologia	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Plataforma tecnológica</b>	Ausência de informação	Selecionar plataforma tecnológica	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Tipo de célula</b>	Ausência de informação	Informar tipo de célula	Completeza	Catálogo de <i>dataset</i> no AgDados

O arquivo do relatório gerado pelo revisor, com apontamentos para a correção da catalogação (Anexo 1), deverá ser anexado junto aos demais arquivos do datasets em revisão, diretamente no AgDados. Ao catalogador caberá analisar, discutir e fazer as correções solicitadas pelo revisor, devolvendo ao sistema o arquivo do relatório por ele corrigido.

## Referências

BALL, A. **Review of data management lifecycle models**. Bath University of Bath, 2012. 15 p.

BRANCO, Z. de S.; MACHADO, D. R.; CESTARI, B. H. P. de S.; OLIVEIRA, Z. P. Controle de qualidade em catalogação cooperativa. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 18., 2014, Belo Horizonte. **Anais...** Belo Horizonte: UFMG, 2014. 21 p.

CÓDIGO de catalogação anglo-americano. 2. ed., rev. 2002. São Paulo: FEBAB: Imprensa Oficial do Estado de São Paulo, 2004. 1 v. Paginação irregular.

DATAVERSE. **Harvard Dataverse**. [Cambridge, MA: Harvard College, 2020]. Disponível em: <<https://dataverse.harvard.edu/>>. Acesso em: 2 maio 2020.

EMBRAPA. **Manual de referência para recursos da informação da Embrapa**. 5. ed. Brasília, DF, 2020. Disponível em: <<https://www.embrapa.br/manual-de-referenciacao>>. Acesso em: 18 out. 2020.

ESTADOS UNIDOS. Department of Agriculture. **NAL Agricultural Thesaurus and Glossary**. Disponível em: <<https://agclass.nal.usda.gov/agt.shtml>>. Acesso em: 8 jul. 2020.

FAO. **AGROVOC**. Rome, 2020. Disponível em: <<http://aims.fao.org/est-registry/vocabularies/agrovoc>>. Acesso em: 8 jul. 2020.

GO FAIR. **FAIR principles**. [2016]. Disponível em: <<https://www.go-fair.org/fair-principles/>>. Acesso em: 27 ago. 2020.

GRAHAM, P. S. Quality in cataloging: making distinctions. **The Journal of Academic Librarianship**, v. 16, n. 4, p. 213-218, Sept.1990.

MORRIS, D. E.; WOOL, G. Cataloging: librarianship's best bargain. **Library Journal**, v. 124, n. 11, p. 44-46, 1999.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION (Estados Unidos). **Taxonomy**. Bethesda. Disponível em: <<https://www.ncbi.nlm.nih.gov/taxonomy>>. Acesso em: 10 jul. 2020.

RIBEIRO, A. M. de C. M. **AACR2**: Anglo-American cataloguing rules, 2nd edition: descrição e pontos de acesso. Brasília, DF: Ed. do Autor, 1995. 577 p.

RIBEIRO, A. M. de C. M. **Catálogo de recursos bibliográficos pela AACR2 2002**: Anglo-American Cataloguing Rules, 2nd edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular.

SINAEPOURFAR, A.; MASIP-BRUIN, X.; GARCIA, J.; MARÍN-TORDERA, E. **A survey on data lifecycle models**: discussions toward the 6Vs challenges. [Barcelona: Universitat Politècnica de Catalunya, 2015]. Technical Report, UPC-DAC-RR-2015-18.

## ANEXO 1 – Modelo de Relatório da Revisão

### Apontamentos para a correção da catalogação de *datasets*

**Projeto:** RmVac – Prática para estruturação de dataverse com subdataverses [exemplo]

**Dataverse:** Solução de inovação 1 do dataverse RmVac – Prática 2 (S1-RmVac-Pratica2) [exemplo]

**Dataset:** <https://doi.org/10.80272/NUT6RT> [exemplo]

#### Metadados de Citação

Elemento (campo/subcampo)	Descrição do erro e/ou inconsistência	Orientação para correção	Elemento(s) de impacto na qualidade	Documento orientador
<b>Título</b>	Ausência de clareza e integridade na atribuição do título.	Atribuir título observando: Tipo de estudo; Amostra estudada; Espécie estudada.	Completeza Clareza Integridade	Catalogação de <i>dataset</i> no AgDados
Subtítulo	Ausência de informação do subtítulo.	Informar subtítulo.	Completeza	Catalogação de <i>dataset</i> no AgDados
Título alternativo	Ausência de informação do título alternativo.	Informar título alternativo.	Completeza	Catalogação de <i>dataset</i> no AgDados
<b>URL alternativa</b>	Ausência da URL alternativa.	Informar URL alternativa válida.	Completeza Precisão	Catalogação de <i>dataset</i> no AgDados
<b>Outro identificador</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Agência	Ausência do nome da agência do outro	Informar nome da agência do outro	Completeza	Catalogação de

	identificador.	identificador.		<i>dataset</i> no AgDados
Identificador	Ausência de identificador digital. Identificador digital não localizado.	Informar identificador digital válido.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Autor</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Nome	Ausência de padronização na descrição do nome.	Seguir regras de catalogação.	Fidelidade a padrões	Catálogo de <i>dataset</i> no AgDados
Afiliação	Ausência de padrão na descrição de nome corporativo.	Seguir regras de catalogação.	Fidelidade a padrões	Catálogo de <i>dataset</i> no AgDados
Sistema identificador	Ausência de indicação de sistema identificador.	Informar nome do sistema identificador.	Completeza	Catálogo de <i>dataset</i> no AgDados
Identificador	Ausência de identificador persistente válido.	Informar nome identificador persistente.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Contato</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Nome	Ausência de padronização na descrição do nome.	Seguir regras de catalogação.	Fidelidade a padrões	Catálogo de <i>dataset</i> no AgDados
Afiliação	Ausência de padrão na descrição de nome corporativo.	Seguir regras de catalogação.	Fidelidade a padrões	Catálogo de <i>dataset</i> no AgDados

E-mail do contato	E-mail incompleto.	Informar e-mail válido.	Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Descrição</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Texto	Ausência de clareza e coerência textual no campo Texto Ausência de precisão na descrição quanto à ortografia, pontuação e transcrição textual na descrição.	Revisar a redação do campo Texto, observando clareza, coerência e precisão quanto à ortografia, pontuação e transcrição textual.	Clareza e integridade. Coerência e consistência	Catálogo de <i>dataset</i> no AgDados
Data	Ausência de informação da Data de descrição.	Informar Data de descrição.	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Assunto</b>	Ausência de área(s) de domínio representativa do assunto.	Adicionar outra(s) área(s) de domínio.	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Palavras-chave</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Termo-chave	Ausência (falta) de termos-chave. Termo-chave escolhido não pertence ao vocabulário escolhido.	Inserir mais termos-chave Corrigir transcrição de termo-chave	Completeza Confiabilidade da informação Coerência e consistência da informação	Catálogo de <i>dataset</i> no AgDados
Vocabulário controlado	Ausência (falta) de indicação de nome(s) do(s) vocabulário(s) controlado(s) usado(s) para o(s) termo(s)-chave	Inserir o nome do vocabulário(s) controlado(s) para usado para indicar o(s) termo(s)-chave	Completeza da informação	Catálogo de <i>dataset</i> no AgDados
URL do termo-chave	Ausência da URL do termo-chave. URL do termo-chave não localizado.	Informar URL válida do termo-chave	Completeza Precisão	Catálogo de <i>dataset</i> no

				AgDados
<b>Classificação</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Termo de classificação	Ausência de termos de classificação. Termo de classificação escolhido não corresponde ao assunto do <i>dataset</i> .	Adicionar mais termos de classificação Eliminar termo de classificação não correspondente ao tema do <i>dataset</i> .	Completeza Confiabilidade da informação Coerência e consistência da informação	Catálogo de <i>dataset</i> no AgDados
Vocabulário controlado da classificação	Ausência de nome(s) de vocabulário(s) controlado(s).	Informar nome(s) de vocabulário(s) controlado(s).	Completeza Precisão Fidelidade a padrões	Catálogo de <i>dataset</i> no AgDados
URL do termo de classificação	Ausência da URL de termo(s) de classificação. URL do termo de classificação não válida.	Adicionar URL de termo(s) de classificação válida Informar URL do termo válida.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Publicação relacionada</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Citação	Ausência de padronização na referência da citação.	Seguir regras de referência bibliográfica.	Fidelidade a padrões	Manual de referência para recursos da informação da Embrapa
Tipo de identificador	Ausência de indicação de tipo de identificador.	Informar tipo de identificador.	Completeza	Catálogo de <i>dataset</i> no AgDados
Número do identificador	Ausência de número de identificador válido.	Informar nome identificador válido.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados

URL da publicação	Ausência da URL da publicação.	Informar URL válida da publicação.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Notas</b>	Ausência de clareza e coerência textual da nota. Ausência de precisão, quanto à ortografia, pontuação no texto da nota.	Revisar a redação de nota, observando clareza, coerência e precisão textual quanto à ortografia, pontuação.	Clareza e integridade Coerência e consistência	Catálogo de <i>dataset</i> no AgDados
<b>Idioma</b>	Ausência de informação sobre o idioma <i>dataset</i> .	Informar o idioma do <i>dataset</i> .	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Produtor</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Nome do produtor	Ausência de padronização na indicação do nome do produtor.	Seguir regras de catalogação.	Fidelidade a padrões	Catálogo de <i>dataset</i> no AgDados
Afiliação do produtor	Ausência de padrão na descrição de nome corporativo.	Seguir regras de catalogação.	Fidelidade a padrões	Catálogo de <i>dataset</i> no AgDados
Abreviatura do nome do produtor	Ausência de abreviatura do nome do produtor.	Informar a abreviatura do nome do produtor.	Fidelidade a padrões	Catálogo de <i>dataset</i> no AgDados
URL do produtor	Ausência da URL do produtor.	Informar URL válida do produtor.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
URL do logotipo do produtor	Ausência da URL do logo do produtor.	Informar URL válida do logo do produtor.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Data de produção</b>	Ausência de data de produção.	Informar data de produção.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados

<b>Local de produção</b>	Ausência de local de produção	Informar local de produção.	Completeza Precisão	Catálogoação de <i>dataset</i> no AgDados
<b>Colaborador</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Tipo de colaborador	Ausência de informação sobre o tipo de colaborador.	Selecionar tipo de colaborador.	Completeza Precisão	Catálogoação de <i>dataset</i> no AgDados
Nome do colaborador	Ausência de padronização na indicação do nome do colaborador.	Seguir regras de catálogoação.	Fidelidade a padrões	Catálogoação de <i>dataset</i> no AgDados
<b>Financiamento</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Nome do órgão financiador	Ausência do nome do órgão financiador	Indicar nome do órgão financiador.	Completeza Precisão	Catálogoação de <i>dataset</i> no AgDados
Número do contrato	Ausência do número do contrato de financiamento.	Informar número de contrato de financiamento.	Completeza Precisão	Catálogoação de <i>dataset</i> no AgDados
<b>Distribuidor</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Nome do distribuidor	Ausência do nome do distribuidor do <i>dataset</i> . Ausência de padronização na indicação do nome do distribuidor.	Indicar nome do distribuidor. Seguir regras de catálogoação na indicação do nome do distribuidor.	Completeza Precisão	Catálogoação de <i>dataset</i> no AgDados
Afiliação do distribuidor	Ausência de padrão na descrição do nome do distribuidor.	Seguir regras de catálogoação.	Fidelidade a padrões	Catálogoação de <i>dataset</i> no AgDados
Abreviatura do nome do distribuidor	Ausência de abreviatura do nome do distribuidor.	Informar a abreviatura do nome do distribuidor.	Fidelidade a padrões	Catálogoação de <i>dataset</i> no AgDados



URL do distribuidor	Ausência da URL do distribuidor.	Informar URL válida do distribuidor.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
URL do logotipo do distribuidor	Ausência da URL do logo do distribuidor.	Informar URL válida do logo do distribuidor.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Data de distribuição</b>	Ausência de data de distribuição.	Informar data de distribuição.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Depositante</b>	Ausência de padronização na indicação do nome do depositante.	Seguir regras de catalogação na indicação do nome depositante.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Data do depósito</b>	Ausência de data de depósito.	Informar data de depósito.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Período da cobertura de dados</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Data de início	Ausência de data de início da cobertura dos dados.	Informar data de início da cobertura dos dados.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
Data de término	Ausência de data de término da cobertura dos dados.	Informar data de término da cobertura dos dados.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Período de coleta de dados</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Data de início	Ausência de data de início da coleta dos dados.	Informar data de início da coleta dos dados.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados

Data de término	Ausência de data de término da coleta dos dados.	Informar data de término da coleta dos dados.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Tipo de dados</b>	Ausência de informação do(s) tipo(s) de dados do <i>dataset</i> .	Informar tipo(s) de dados do <i>dataset</i> .	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
<b>Série</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Nome da série	Ausência do nome da série do <i>dataset</i> .	Informar nome da série, caso exista.	Completeza Precisão	Catálogo de <i>dataset</i> no AgDados
Informação sobre série	Ausência de informações sobre a série do <i>dataset</i> . Ausência de clareza e coerência textual na informação sobre a série.	Fornecer informações sucintas sobre a série, caso exista. Revisar redação da informação sobre a série.	Clareza e integridade Coerência e consistência	Catálogo de <i>dataset</i> no AgDados
<b>Software</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
Nome de software	Ausência do nome do software utilizado.	Indicar o nome do software	Completeza Confiabilidade das informações	Catálogo de <i>dataset</i> no AgDados
Versão do software	Ausência da informação sobre a versão do software utilizado.	Indicar o nome do software.	Completeza Confiabilidade das informações	Catálogo de <i>dataset</i> no AgDados
<b>Material relacionado</b>	Ausência de informação sobre material relacionado.	Informar material utilizado relacionado ao <i>dataset</i> , caso exista.	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Datasets relacionados</b>	Ausência de informação de <i>datasets</i> relacionados.	Informar <i>dataset</i> relacionados, caso existam.	Completeza	Catálogo de <i>dataset</i> no

				AgDados
<b>Outras referências</b>	Ausência de informação sobre outras referências ao <i>dataset</i> .	Informar outras referências relacionadas ao <i>dataset</i> , caso existam.	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Fontes de dados</b>	Ausência de informação sobre fontes de dados do <i>dataset</i> .	Informar fontes de dados do <i>dataset</i> , caso existam.	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Origem das fontes de dados</b>	Ausência de informação sobre origem das fontes dos dados do <i>datasets</i> .	Informar origem das fontes de dados do <i>dataset</i> , caso exista.	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Características das fontes de dados observadas</b>	Ausência de informação sobre características das fontes de dados observadas.	Informar sobre características das fontes de dados observadas, caso existam.	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Documentação e acesso a fontes de dados</b>	Ausência de informação sobre documentação e acesso a fontes de dados do <i>dataset</i> .	Informar sobre documentação e acesso a fontes de dados do <i>dataset</i> , caso exista.	Completeza	Catálogo de <i>dataset</i> no AgDados

## Metadados *Life Science*

Elemento (campo)	Descrição do erro e/ou inconsistência	Orientação para correção	Elemento(s) de impacto na qualidade	Documento orientador
<b>Tipo de design</b>	Ausência de informação	Selecionar tipo de design	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Tipo de fator</b>	Ausência de informação	Selecionar tipo de fator	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Organismo</b>	Ausência de informação	Selecionar organismo	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Outro organismo</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
<b>Tipo de medida</b>	Ausência de informação	Selecionar tipo de medida	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Outro tipo de medida</b>	Não se aplica	Não se aplica	Não se aplica	Não se aplica
<b>Tipo de tecnologia</b>	Ausência de informação	Selecionar tipo de tecnologia	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Plataforma tecnológica</b>	Ausência de informação	Selecionar plataforma tecnológica	Completeza	Catálogo de <i>dataset</i> no AgDados
<b>Tipo de célula</b>	Ausência de informação	Informar tipo de célula	Completeza	Catálogo de <i>dataset</i> no AgDados



*Informática Agropecuária*

